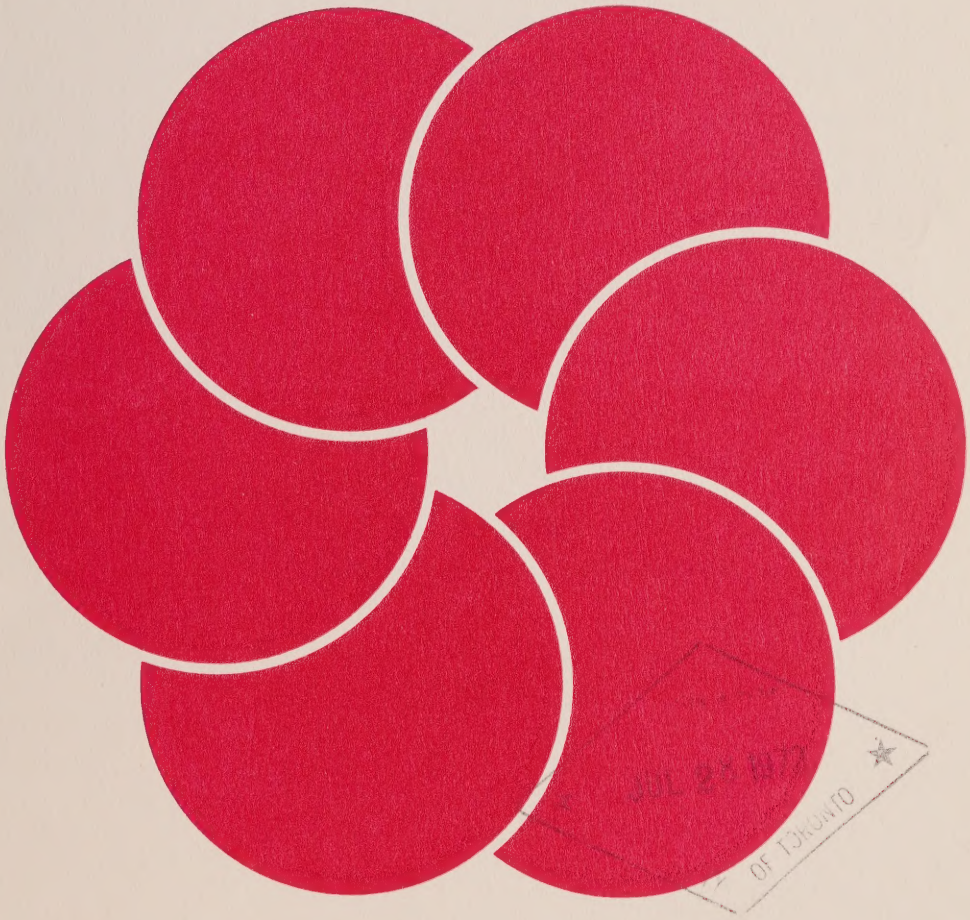
CATALOGUE No. 13-537 OCCASIONAL

Econometric Study of Incomes of Canadian Families

1967



Canada

DOMINION BUREAU OF STATISTICS

General Publicati

DOMINION BUREAU OF STATISTICS
Consumer Finance Research Staff

ECONOMETRIC STUDY OF INCOMES OF CANADIAN FAMILIES

1967

Published by Authority of
The Minister of Industry, Trade and Commerce

June 1972
1100-513

Price: 75 cents

Information Canada
Ottawa

Publications Dealing with INCOMES

Catalogue	Title
13 - 525	Incomes, Assets and Indebtedness of Non-farm Families in Canada, 1963, O., E.
13 - 528	Income Distributions by Size in Canada, 1965, O., E.
13 - 529	Incomes of Non-farm Families and Individuals in Canada, Selected Years 1951 - 65, O., E.
13 - 534	Income Distributions by Size in Canada, 1967, O., E. and F.
13 - 535	Earnings and Work Experience of 1967 Labour Force, O., E.
13 - 536	Statistics on Low Income in Canada, 1967, O., E.
13 - 538	Family Incomes (Census Families), 1967, O., E.
13 - 539	Comparative Income Distributions, 1965 . 67, O., E.
13 - 540	Household Facilities by Income and Other Characteristics, 1968, O., E.
13 - 544	Income Distributions by Size in Canada, 1969, O., Bil.
99 - 544	Incomes of Canadians by Jenny R. Podoluk, 1961 Census Monograph, O., E.

O. — Occasional E. — English F. — French Bil. — Bilingual

In addition to the selected publications listed above, Statistics Canada publishes a wide range of statistical reports on Canadian economic and social affairs. A comprehensive catalogue of all current publications is available free on request from Statistics Canada, Ottawa, Ontario K1A 0T6.

The Department of National Revenue publishes annually "Taxation Statistics, Part 1 — Individuals", which may be obtained from Information Canada, Ottawa.

PREFACE

This report is one of a number of special reports prepared from the data collected by the Survey of Consumer Finances in the spring of 1968. It is an analytical study of the determinants of family income in 1967 for the five regions and Canada. The study ranks socio-economic characteristics of families in terms of their influence on family incomes. It attempts to quantify the effect of the different variables on family income and comments on inter-regional differences in this respect. The study uses a multiple regression technique for estimating these effects.

Mr. R.K. Chawla from the Consumer Finance Research Staff planned and executed the study. He also wrote the report under the direction of Mrs. G. Oja. Mr. J. Lewis from the Regional Research Staff provided advice and assistance on computer programming aspects of the study.

WALTER E. DUFFETT,
Chief Statistician of Canada.

SYMBOLS

The following standard symbols are used in Dominion Bureau of Statistics publications:

.. figures not available.

... figures not appropriate or not applicable.

— nil or zero.

-- amount too small to be expressed.

p preliminary figures.

r revised figures.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Several economists and other social scientists have dwelt upon the influence of socio-demographic and economic characteristics on either the family's income or the individual's earnings.[1] Their major concerns have been to demonstrate the pertinence of the individual's age, education, occupation and sex on the variations in income taken in terms of either grouped data or a percentile presentation. Hardly any of these researchers has attempted to demonstrate the significance of the given socio-demographic or economic characteristics in a systematic way or ranked them in terms of their importance; e.g., whether the education, age or sex of an individual was equally effective in relation to his income and, if not, which of the characteristics was most important, second important and so on. In other words, one could not simply postulate an income model in relation to the given sample characteristics and arrive at a conclusion that all these characteristics were equally effective or could be measured on the same scale.

Thus the primary objective of this study is to present an analytical procedure which would list the given socio-demographic characteristics affecting income in a preferential order and also quantify their individual effects. The income concept under study is that of the economic family which by definition, is a group of all relatives living together in the same household and related by blood, marriage or adoption. Thus all relatives living together comprised one family unit whatever the degree of family relationship. The present exposition does not analyse the incomes of unattached individuals which by definition, are persons living by themselves or living in a household where they are not related to other members of the household.

The income variable in this study is "total family income" which is a sum of (i) wages and salaries before tax deductions; (ii) military pay and allowances; (iii) net income from non-farm self-employment; (iv) net income from farm self-employment; (v) income from roomers and boarders; (vi) interest and dividends; (vii) other income from investments; (viii) family and youth allowances; (ix) old age pensions; (x) unemployment insurance benefits; (xi) other income from government sources; (xii) retirement pensions and (xiii) other money income. This income variable is analyzed by a multiple regression technique using extensively dummy variables representing the non-measurable or the qualitative socio-demographic characteristics of an economic family. The resulting empirical analysis has been presented both at the regional and the national levels.

[1] See Bjerke (6) for an interesting and a comprehensive survey of the literature on the income and wage distributions.

In sum, our objectives in this paper are:

(a) to quantify the effects of each of the socio-demographic characteristics used in the postulated income model (i.e., by using the regression coefficients),

(b) to quantify the individual contributions of such characteristics to the total explained variance of income both at the regional and the national levels (by using the analysis of variance technique), and

(c) to describe the cardinal rankings of these characteristics.

Let us now outline briefly the contents of other Chapters of this exposition. Chapter 2 describes the survey data on which the present econometric analysis is based. This includes: (i) the size of the sample and its reference period; (ii) the weighting factor used; (iii) the nature of the questionnaire used and finally (iv) the background of the socio-demographic characteristics selected for the study.

Chapter 3 explains the statistical methodology employed, the postulation of the model and the selection of cross-classified characteristics included. It defines the basic sets of hypotheses and the interpretation of the estimated coefficients.

Chapter 4 contains the empirical analysis at the regional level. It focuses attention on the cardinal rankings of explanatory socio-demographic characteristics for the individual regions, their quantitative contributions and the relative differences in incomes with respect to these characteristics of economic families (using analysis of variance and the regression coefficients). Finally, it takes into account the question of goodness of fit of the models and the related tests of significance.

Chapter 5 is almost a replica of Chapter 4 except that it focuses attention at the all Canada level. In this situation we have a slightly extended version of the income model which contains two additional socio-demographic characteristics as explained in Chapter 2.

Chapter 6, as an epilogue to the text, summarizes the major conclusions drawn from such econometric models, their limitations in view of the operational, computational and other existing resource constraints. The chapter also discusses the possible extensions and improvements in the model which could not be incorporated in this study.

There are three appendices to the text. Appendix A presents a complete list of both the qualitative and quantitative variables used in this study. Appendix B summarizes some of the major problems encountered in the process of specifying the income model described here. Beginning with a formal presentation of the initial list of socio-demographic characteristics chosen for this study, the appendix outlines certain technical and methodological considerations which led us to discard or modify some of these characteristics for the final version of the model. Lastly, Appendix C contains the regression results for both the regional and the national models.

CHAPTER 2

DESCRIPTION OF DATA AND SELECTED VARIABLES

The present chapter outlines the nature and background of the statistical data on which the empirical analysis presented in Chapters 4 and 5 are based. In this endeavour, we shall also briefly outline the nature of various characteristics obtained from the questionnaires and our reasons of selecting only those relevant to us.

Statistics Canada (formerly the Dominion Bureau of Statistics) has been conducting surveys on incomes of Canadian families and individuals on a periodic basis since 1952. The major statistics so collected have been total family income[1] by its components along with relevant socio-demographic characteristics such as the sex of the head and other members, their ages, education, marital status, home ownership, the area of residence and regional locations, etc. Since 1965, these surveys have included a representative sample of all private households (with minor exceptions) whereas in the earlier years only non-farm households were surveyed.[2] Further, most of these surveys have been carried out in conjunction with the labour force surveys and a number of other social and behavioural characteristics of all those over 14 years of age, participating or not participating in the labour force have been available. For 1967, a number of statistics pertaining to annual work histories of all those included in the sample were collected.

Thus the present micro-data, (both at the family and the individual level) were collected for the Consumer Finance Research Staff in April 1968. Complete information about the survey, its sampling methodology, the related response rates and various other characteristics have been fully explained elsewhere.[3] The sample consisted of 18,143 economic families which when adjusted by the weighting factor resulted in an estimate of 4.52 million families in Canada. The corresponding regional distribution of these is shown below:

TABLE 2.1. Distribution of Economic Families by Regions

Region	Unweighted sample size		Weighted estimates	
	Actual	% of the total	Actual	% of the total
Atlantic	3,965	21.85	391,090	8.67
Quebec	3,733	20.58	1,245,330	27.57
Ontario	4,866	26.82	1,660,850	36.77
Prairies	3,721	20.51	761,050	16.85
British Columbia	1,858	10.24	459,000	10.16
Canada[1]	**18,143**	**100.00**	**4,517,320**	**100.00**

[1] Excluding the Yukon and the Northwest Territories.

The table conveys an overall impression about the distribution of the sampled and the estimated families over the regions. The former describes the order of degrees of freedom employed in carrying out certain statistical tests as described in Chapter 4 at the regional level; whereas the results of empirical analysis presented in the text apply to the estimated number of families. Sample records have been weighted up to national totals (i.e. by ratio estimation) by using weighting factors which compensated for differential sampling ratios[4] and varying response rates.[5]

---

[1] Total family income is different from the concept of personal income as defined in the system of National Accounts. The latter constitutes income of the personal sector of the economy and includes imputed and other non-cash components.

[2] Data from these surveys have been published in various DBS reports, the latest of them being DBS Catalogue 13-544, *Income Distribution by Size in Canada, 1969.*

[3] For a more comprehensive background, see pp. 7-9, 14-16, and 66-71 of *Income Distributions by Size in Canada, 1967*, Catalogue 13-534, Dominion Bureau of Statistics, Ottawa.

[4] A detailed account of these ratios and their estimation procedures is given in *Methodology: Canadian Labour Force Survey*, Catalogue 71-504, Dominion Bureau of Statistics, Ottawa.

[5] See pp. 66-69, *Income Distributions by Size in Canada, 1967*, Catalogue 13-534, Dominion Bureau of Statistics, Ottawa.

It may, however, be emphasized that the weighting factor is not only pertinent in deriving the estimates of families in the population but also adjusts the weight of different socio-demographic groups in the total population. For instance, in the absence of this factor, each sampled family would be assigned the same socio-demographic significance or a unit value in our model. The following table demonstrates the significance of the weighting factor on some of the socio-demographic characteristics of a family:

TABLE 2.2. Sample Correlations Between Some Selected Characteristics and the Weighting Factor[1]

Characteristic[2]	Correlation coefficient	Characteristic[2]	Correlation coefficient
AR_1	0.2289	EH_1	– 0.0695
AR_2	– 0.0139	EH_2	– 0.1840
AR_3	– 0.0735	EH_3	– 0.0973
AR_4	– 0.2092	EH_4	0.0036
SH_1	0.1172	EH_5	0.0901
SH_2	– 0.1172	EH_6	0.0738
		EH_7	0.2811

[1] These sample correlations have been evaluated from 1/3rd of the total sample at the national level. The values (in absolute terms) greater than 0.0252 are significant at 5% level.
[2] For the definitions of characteristics, see Appendix A.

Differences in the correlation coefficients in the above table indicate that weighted and unweighted estimates by the categories of the above selected characteristics would differ significantly. The characteristics were picked as examples to demonstrate the importance of using a weighted regression technique. An unweighted regression, on the other hand, would have resulted in biased and inefficient estimates.

In view of our objective to explain the total annual income of a family, we selected from the sample questionnaires[6] only those socio-demographic characteristics which were valid and meaningful on an annual basis. These included (i) regional location; (ii) area of residence; (iii) sex of head; (iv) marital status of head; (v) education of head; (vi) immigration status of head; (vii) main occupation of head; (viii) the number of weeks head worked during the year; (ix) education of wife; (x) the number of weeks wife worked during the year; (xi) age of head; (xii) family size and (xiii) the number of earners in a family. The econometric model so derived has been specified from this core of thirteen basic characteristics.

It must be stressed here that not all of the above listed characteristics directly define the final version of the model used for empirical inferences. There are certain characteristics which needed some form of modification and there are some which were used to create new characteristics. A brief account of these developments has been given in Appendix A. But for the present purpose we may, however, mention that a final list of socio-demographic characteristics defining a regional income model included (i) area of residence; (ii) education of head; (iii) occupation of head in conjunction with his education; (iv) wife's participation in labour force in conjunction with her education; (v) head's participation in labour force; (vi) age of head; (vii) the proportion of earners in a family obtained by the ratio of the number of earners to the total size of family, and (viii) the type of family, i.e., whether a "husband-wife" family or "other". The latter included all families with female heads and those with male heads with marital status either single, divorced or legally separated, etc. On the other hand, the model at the all Canada level included, besides all characteristics of the regional model, two additional socio-demographic attributes namely (ix) the regional location of family and (x) the immigration status of its head. The omission of item (x) from the regional model is due to certain conceptual and operational problems, whereas the exclusion of item (ix) in this situation is self-explanatory.

The above list of exogenous socio-demographic characteristics used in the income model is partly based on what was available and partly what other researchers in this field have established. Some of the questions which naturally follow this simple specification of characteristics is in what way do

[6] See the formats of these questionnaires on pp. 80-83, *Income Distributions by Size in Canada, 1967*, Catalogue 13-534, Dominion Bureau of Statistics, Ottawa.

they affect the family's income and what are their individual contributions in explaining the income variability? The remaining paragraphs of this chapter deal with the first question as to how these characteristics influence income of a family whereas the second question about their individual contributions and other quantitative effects will be dealt with in Chapters 4 and 5. Thus a brief description of how the selected characteristics affect income is presented below in the same order in which they have been listed under items (i) to (ix).

(i) Area of Residence

The nature of residential location of a family, quite often conveys a fair impression about its characteristics. For example, certain occupations and income levels predominate in metropolitan areas whereas in rural areas other occupations such as farmers and farm workers are prevalent.[7]

(ii) Education of the Head of Family

Like other economists,[8] we have treated education as a key indicator of one's earnings potential. Using the most conventional terminology, all of these economists have regarded education as a proxy for investment in human capital, the level of which is determined mainly by the total number of years spent in schooling and the corresponding rates of return in terms of the future flow of earnings. Family incomes are heavily dependent on the head's earnings which in turn are largely determined by his level of education.

(iii) Occupation of the Head of Family in Conjunction with His Education

For our purposes, we define the significance of occupation as a link between one's education and income; in other words, education enables an individual to participate in the labour force or other productive activities, the pursuance of which yields him a return, i.e., income, depending upon the quality and quantity of his efforts. According to Tinbergen (57), each individual, facing a labour market offering different sets of prices for different categories of jobs, opts for one which maximizes his utility. The author contended that "this utility will depend on at least three variables... (a) the attributes of the person we consider; (b) the attributes of the job chosen, and (c) the income so to be earned...." (p. 494). We shall mainly be dealing with item (c) with a slight modification that our analysis depicts differences in incomes of families resulting from different jobs held by their heads.

Confronted with a choice of defining an individual's activity, we had the option whether to categorize the individual by the type of function he performs (i.e. occupation) or by the nature of economic activity he is associated with (i.e. industry)—in other words, defining occupation in terms of broad occupational groups such as the managerial, professional and technical, etc. or by sectors such as agriculture, manufacturing, etc. In our judgement, occupations are more homogeneous as to income than industries. No matter what definition one chooses, one cannot turn a blind eye to the methodological and interpretative complexities created by large scale grouping, which in turn, may influence the empirical inferences.[9]

For the present study, data for nine broad occupational groups[10] based on the standard occupational classification[11] were available. Then to account for the income differences within each of such groups, (in other words, minimizing the effects of grouping individuals with heterogeneous characteristics) we have cross-classified these broad groups by levels of education.[12]

Another word of caution may, however, be added about the limitation of this technique too; i.e., it fails to take into account the income differences arising from the nature of duties one performs, or one's position in the hierarchal set-up or the degree of responsibility, etc., as it is quite plausible to find two people with similar education and professional class earning significantly different incomes. Lydall (34) has pointed out that individuals in the higher income brackets are not paid according to their abilities measured by variable such as formal education, but according to their responsibility as leaders.[13]

(iv) Wife's Participation in Labour Force in Conjunction with Her Education

The wife's decision to undertake a job and supplement the family's income is primarily influenced by (i) socio-demographic and (ii) economic factors. The former includes her (a) age, (b) education, (c) the number of years married, (d) the number of children by age groups, (e) the region and area of residence and (f) the family size

---

[7] Definitions of metropolitan, urban and rural areas are given in Appendix A.

[8] Mincer (37,39), Thurow (56), Hansen (27), Becker and Chiswick (4), Hanoch (26), Tinbergen (57), Lydall (34), Chiswick (13,14) and Podoluk (46).

[9] A possible solution to overcome this formidable problem of grouping would have been to stratify data into a varying number of occupational groups and develop models explaining income within each of these groups and then study their differentials. But the nature of sample questionnaire and the existing resource constraints have so far hindered its accomplishment.

[10] Blau and Duncan (7) have described the role of such an occupational stratification in terms of a "link between the economy and the family, through which the economy affects the family's status and the family supplies manpower to the economy" p. 7; for another interpretation, see p. 24.

[11] In terms of the classification used in 1961 Census.

[12] See Chapter 3 and Appendix B for a detailed methodology adopted in listing the cross-classified variables.

[13] Mincer (37) arrived at a somewhat similar statement regarding one's income, training, starting position in one's career and the overall span of working life, see p. 285.

(i.e. other relatives in the household). The economic factors include (g) her previous labour force experience, (h) her husband's earning potential, (i) his occupation, (j) the wage rates for females, (k) the prevailing activity rates of females by age groups partly conditioned by the demand for female workers in different sectors of the economy, (l) the overall unemployment rates for both males and females, (m) the incidence of husband's unemployment, (n) the ownership of home, and (o) the financial liabilities such as debts and mortgages. Besides these characteristics, there are certain other factors such as boredom caused by being restricted to house work and a desire to maintain an active and social link with external events and society in general.

An analysis of these socio-demographic and economic characteristics can be found in recent literature where several contributors[14] have demonstrated empirically the corresponding functional dependence of married women's participation in the labour force for both the U.S.A. and Canada. It should be noted that none of the mentioned studies included in full the above listed characteristics. The most common attributes used have been (a), (b), (e), (f), (g), (h), (l) and in some recent studies,[15] (n) and (o) as well. As Cain (10) has pointed out, the demographic factors alone account for very little compared to the purely economic ones in explaining the rising number of married women joining the labour market.

In this study, we have considered the wife's participation during the year in the labour force in relation to her education[16] primarily because the latter could be used as a proxy for tastes for market work. Like the previous cross-classified group (i.e., head's occupation in conjunction with his education), this group too contained nine explicit cross-classification categories specifying full-year and part-year participation of wives according to their educational levels varying between some elementary education to graduation from university.

(v) Head's Participation in Labour Force

We measured the effect of the number of weeks worked by head in terms of two broad categories i.e., either the head worked between 1 - 52 weeks a year or he did not work at all. The corresponding hypothesis being that the economic returns to the former group would be greater compared to the latter group. Such categorization of the number of weeks worked is bound to create certain problems, especially those emerging from the process of broad grouping. For example, this form of categorization does not account for the differences in incomes accruing to jobs affected by seasonal factors. It is not that we were unaware of these problems of grouping, but the presence of technical problems made it necessary to choose this form of specification. Maybe it would be appropriate to point out here that the empirical estimates pertaining to this classification could be weak because of the bias resulting from such heterogeneous grouping.

(vi) Age of the Head of Family

The specification of age variable in an income model does not merely imply that income of an individual is related to his or her age taken in terms of actual years. In fact, the upward sweep in the earning curve implies the effects of experience gained, skills acquired, and other personal improvements in terms of physical strength and motivation.

For analytical purposes, we have used the age variable in terms of actual number of years and its corresponding squared term in the model in order to express the relationship between the family's income and the head's age as a parabolic function of second degree.[17] This in turn enabled us to locate the specific age when the income reached its peak beyond which the income curve started declining.

(vii) Proportion of Earners in a Family

In view of the fact that we are considering the income of an economic family, we must take into account the economic contribution of all members (related by blood, marriage or adoption) living in the same unit. The economic contribution in this instance is their relative shares in the aggregate income of the unit. As we are predicating our judgements on the basis of a single equation model explaining total income, we could not sub-divide such individual contributions by type of earners and their related socio-demographic characteristics. Therefore, what we propose to explain here is the effect of the variable NE/FS, i.e., the ratio of the number of earners to family size, on the family's total income. Like the age variable, this too was expressed as quantitative in the model.

(viii) Type of Family

The present empirical analysis describes differences in the incomes of husband and wife families in contrast to single parent families and all other family types. The necessity to include this variable in the model arose from two considerations. First, a

---

[14] Ostry (44,45) and Spencer and Featherstone (52) have dealt with the economics of labour force participation of women in Canada. In terms of methodology, all of these studies have used dummy variables representing the qualitative attributes in question and the simple least squares technique for estimation.

[15] Rosett (19) and Spencer and Featherstone (52).

[16] It may be mentioned in passing that the levels of the husband's and wife's education showed remarkably high correlation coefficients.

[17] See Cain (10), Hanoch (26), Bowen and Finnegan (8) who have used the age variable in grouped form and finally expressed it in terms of dummy variables in their econometric studies.

meaningful specification of the exogenous characteristic pertaining to the wife's participation in labour force required such treatment as not all of the economic families had wives present. Secondly, to maintain conformity with our recently published report[18] describing differences in incomes by types of families, we wished to examine them further in the light of several other socio-demographic characteristics.

(ix) Regional Location of Family

An inter-regional comparison of income models, each containing the above listed eight characteristics, would fairly explain the income differences ascribing to these characteristics of families living in different regions of Canada. But the question is how to evaluate the regional effects on the income of a family in the light of its other socio-demographic characteristics at the national level. To resolve this particular problem, we developed a national model specifying the regional locations, namely the Atlantic, Quebec, Ontario, the Prairies and British Columbia. Regional differences in economic conditions and incomes have been treated in the Canadian literature quite extensively.[19] It is generally recognized as being one of the most urgent concerns for economic policy.

[18] See Tables 13 and 14 in *Income Distributions by Size in Canada, 1967*, Catalogue 13-534, Dominion Bureau of Statistics, Ottawa.

[19] For instance, see Brewis (9), Chernick (12), Mann (35), Nicholson and Sametz (42), Putnam & Putnam (47), Sitwell (50), Stone (54) and Vallee (58).

(x) Immigration Status of the Head of Family

Immigrants constitute an important segment of Canada's population. Their economic status and length of stay in the country are inter-related in a very complex way. In this study, we have classified families by the immigration status of their heads taken in terms of the year they arrived in Canada. Further, for analytical purposes, such a status of family head was stratified into four groups, namely, (i) born in Canada, (ii) arrived before 1946, (iii) arrived between 1946 to 1966 and finally, (iv) those whose status was not ascertained.

Such an exposition of socio-demographic characteristics used in the model would be incomplete without a mention of those characteristics which influence the family's income but have not been treated explicitly in this study because relevant data do not exist. These included the head's (as well as other family members) intelligence, endurance to work, willingness to take risks and other challenging tasks, mobility, family background, responsiveness to incentives and other attitudes, etc. Similarly, information is lacking about the wife's other activities, tastes about work, etc., which influence her contribution to the family's income. In sum, the empirical analysis both at the regional and the national levels, presented in Chapters 4 and 5, have been confined to the effects of well-known socio-demographic characteristics on the family's income. The methodology adopted in this respect is the subject of the next chapter.

CHAPTER 3

A STATISTICAL METHODOLOGY OF THE MODEL

This chapter explains the statistical methodology adopted for developing the income models both at the regional and the national levels. In this perusal we shall also define the sets of hypotheses pertaining to the exogenous socio-demographic characteristics discussed in Chapter 2. Such sets of hypotheses will define our expectations as to the results of the models.

The basic techniques of quantitative research available to a researcher engaged in analysing income data, are limited. It is not long ago since various economists initiated studies on the measurement and concentration of income of households or communities. As their objectives were simply limited to the measurement, degree of variability and concentration of income, the statistical concepts they used were the standard deviation, coefficient of skewness, Gini coefficient, Pareto and lognormal distributions. But with the development of statistical techniques and the growing concern of researchers to make the best use of the data collected, analytical and empirical research in this field has turned to the use of multivariate techniques.

The adoption of any one of the multivariate techniques[1] depends upon (a) the researcher's objectives, (b) the nature of information available, and (c) the computational, operational and other constraints. Out of these three criteria, the researcher's objectives play the most crucial role in determining the appropriate statistical methodology. In other words, the dimension and complexity of the objectives determine the basic methodology to be used for the purposes of empirical inferences.

In view of our multiple objectives, we opted for a multivariate regression technique. This will enable us to meet the objectives specified in the Introduction.[2] Other important features of multivariate regression analysis are as follows:

(a) its' computational flexibility enables one to drop the insignificant exogenous attribute and re-estimate the model without excessive computation;

(b) its' estimation procedure enables one to study the behaviour of exogenous attributes, i.e., whether they are interdependent, or statistically relevant to the dependent variable.[3]

The present regression model contains only three quantitative variables, namely (i) the endogenous variable which is the total family income, (ii) the age of head (in actual years) and (iii) the proportion of earners in a family. The remainder of the exogenous variables have been expressed in terms of the dummy variables representing the qualitative attributes of a family such as its regional location, area of residence, etc. It is basically a device of expressing qualitative data in quantitative terms and thus making it possible to subject it to the usual methods of estimation, testing of significance, etc.

There are certain advantages and disadvantages of using dummy variables. In terms of advantages, first, their estimated coefficients reveal the differences between the effects of the given variables and their corresponding reference variables which have to be selected before the actual estimation takes place, otherwise the equation could not be estimated.[4] Secondly, the statistical significance of these variables, tested by conventional t-test, reveal the significance of the differences between the given variables and their reference variables. On the other hand, the major disadvantage of using dummy variables is that they restrict the researcher's manoeuvreability to assign any other form of functional relationship *except* additivity between the given variables. Of course, one can apply a logarithmic or reciprocal transformation to the endogenous variable, but as far as the exogenous variables are concerned, there is hardly any choice in specification. Thus this form of forced additivity is bound to create certain odd results in the form of either wrong signs associated with the estimated coefficients or their lack of significance.[5]

A salient feature of the income model described here is that it embodies certain socio-demographic characteristics in terms of cross-classification within its basic framework. In other words, we did not have to split the data into any form of subgroups in order to study the effects of characteristics taken two at a time, like some researchers in in this field have done.[6] Moreover, the advantages of incorporating characteristics in terms of cross-classification within the same framework of the model are, first, one can study the differential impact of one cross-classified group keeping the others fixed and secondly, it reduces the computational costs.

---

[1] See Kelly, Beggs and McNeil (30) for an interesting summary of multivariate techniques. Especially refer to their Table 7.4, p. 246.

[2] See points (a) to (c) on p. 7.

[3] The recently developed computer programmes for multiple regressions normally give a simple correlation matrix of exogenous attributes from which one can study their interdependence. It is vital that such an interdependence should be kept at a minimum, otherwise the so-called multicollinearity effects would create various other statistical problems. Although we have not provided these lengthy tables in the present text, the readers wishing to pursue research on these socio-demographic characteristics (either for individual region or for all Canada) can obtain the correlation matrices by writing to Mr. R. Chawla, Consumer Finance Research Staff, Statistics Canada, Ottawa.

[4] For a detailed theory on the use of dummy variables, see Suits (55), Melichar (36) and Goldberger (23), where the latter has presented an interesting mathematical exposition.

[5] We did, however, run a pilot semi-logarithmic model based on 1% of the total sample, where the logarithm of income was regressed on some selected characteristics. The net result was rather disappointing. This is another reason why we used the additive model only.

[6] See Hanoch (26) and Bowen and Finnegan (8), who had to aggregate data in specific groups in order to study the effects of two or more characteristics.

The characteristics chosen in this context were the main occupation of head in conjunction with his education and the wife's participation in labour force in conjunction with her education.[7] The variables pertaining to these cross-classified characteristics were specified in the final version of the model on *a-priori* grounds. In other words, they were treated like all other exogenous variables.

For empirical purposes, the numerous combinations emerging from the cross-classification of these paired characteristics were reduced in order to avoid their over-riding effects on other exogenous variables. In this manner, the characteristic pertaining to the head's occupation in relation to his education constituted 28 variables (including the reference variable) and the one defining the wife's participation in labour force in conjunction with her education contained 9 variables (including the reference variable). The corresponding reference categories included were (i) the heads who did not work during 1967 and had no education and (ii) the wives who did not work during 1967 and had no education.

As we have stated before, the model analyzes the total income of an economic family, which by definition, may or may not have a wife as its member. Therefore, the situation called for some changes so that the end product of the model would explain the income effects (pertaining to the wife's participation in labour force in conjunction with her education) in the case of husband-wife families only. The changes made in this context evolved a new variable differentiating the type of family; i.e., whether it was a husband-wife family or not. The latter group included a single parent or other types of families. A detailed account of such variables and the necessary steps adopted to create them have been provided in Appendix A.

Thus a complete income model so derived at the all Canada level can be postulated as follows:

$$
\begin{aligned}
Y = {} & b_{0,0} + b_{1,1}RG_1 + b_{1,2}RG_2 + b_{1,3}RG_3 + b_{1,4}RG_4 + b_{1,5}RG_5 \\
& + b_{2,1}AR_1 + b_{2,2}AR_2 + b_{2,3}AR_3 + b_{2,4}AR_4 \\
& + b_{3,1}EH_1 + b_{3,2}EH_2 + b_{3,3}EH_3 + b_{3,4}EH_4 + b_{3,5}EH_5 + b_{3,6}EH_6 + b_{3,7}EH_7 \\
& + b_{4,1}ISH_1 + b_{4,2}ISH_2 + b_{4,3}ISH_3 + b_{4,4}ISH_4 \\
& + b_{5,1}MOH(R) + b_{5,2}MOH_{21} + b_{5,3}MOH_{22} + b_{5,4}MOH_{23} + b_{5,5}MOH_{24} \\
& + b_{5,6}MOH_{31} + b_{5,7}MOH_{32} + b_{5,8}MOH_{33} + b_{5,9}MOH_{34} \\
& + b_{5,10}MOH_{41} + b_{5,11}MOH_{42} + b_{5,12}MOH_{43} \\
& + b_{5,13}MOH_{51} + b_{5,14}MOH_{52} + b_{5,15}MOH_{53} \\
& + b_{5,16}MOH_{61} + b_{5,17}MOH_{62} + b_{5,18}MOH_{63} \\
& + b_{5,19}MOH_{71} + b_{5,20}MOH_{72} + b_{5,21}MOH_{73} + b_{5,22}MOH_{81} + b_{5,23}MOH_{82} \\
& + b_{5,24}MOH_{91} + b_{5,25}MOH_{92} + b_{5,26}MOH_{93} + b_{5,27}MOH_{101} + b_{5,28}MOH_{102} \\
& + b_{6,1}WLF+(R) + b_{6,2}WLF+_{21} + b_{6,3}WLF+_{22} + b_{6,4}WLF+_{23} + b_{6,5}WLF+_{24} \\
& + b_{6,6}WLF+_{31} + b_{6,7}WLF+_{32} + b_{6,8}WLF+_{33} + b_{6,9}WLF+_{34} \\
& + b_{7,1}HLFE + b_{7,2}HLFU + b_{8,1}AHF_1 + b_{8,2}AHF_2 \\
& + b_{9,1}NE/FS + b_{10,1}NF_1 + b_{10,2}NF_2 + u \qquad \ldots (3.1)
\end{aligned}
$$

Where Y denotes the total family income, $b_{0,0}$, the constant term and $b_{1,1}$, $b_{1,2}$, . . ., $b_{10,2}$ are the unknown regression coefficients associated with the exogenous variables as defined in Appendix A. It is further assumed that the residual component u of the model satisfies the well known assumptions necessary for carrying out the ordinary least squares estimation.

Thus the all Canada model contains 65 variables (including the constant term) in all. As the application of least squares technique required the omission of one variable from each of the groups of dummy variables, we selected RG_1, AR_4, EH_1, ISH_3, MOH(R), WLF+(R), HLFU and NF_1. In other words, these have been treated as the reference variables throughout the analysis and their corresponding regression coefficients are being equated to zero.[8] So the actual estimated equation at the national level contains 57 variables (including the constant term).

The basic sets of hypotheses pertaining to family income and based on 56 estimated regression coefficients at the national level have been defined in Column 5 of Table 3.1. These inequalities define our expectations as to the order of magnitudes of incomes pertaining to the given socio-demographic attributes. If the numeric order of the estimated regression coefficients corresponding to say, the i-th characteristic, followed our *a-priori* specification of b's corresponding to the i-th characteristic, we would then conclude that the stated hypothesis

[7] Appendix B describes the background to the selection of these characteristics.

[8] It may be mentioned that the selection of reference variables is arbitrary and does not affect the final estimates obtained from the model, that is based on dummy variables.

in the case of the i-th characteristic was fully satisfied. Similarly if half or more than half of the *a-priori* inequalities within a set were met in terms of the numeric magnitudes of estimated regression coefficients corresponding to the n-th characteristic, we would conclude that the stated hypothesis pertaining to the n-th characteristic was partly satisfied. An abbreviated interpretation of each of the hypotheses is given in Column 6 of Table 3.1.

TABLE 3.1. Sets of Hypotheses Pertaining to Family Income and Their Interpretation

Characteristic	Symbol[1]	Number of variables within characteristic (excluding reference variable)	Hypothesis pertaining to each characteristic	Interpretation of hypothesis (taking into account the extreme inequalities as shown in the preceding column)
				Family income:
1. Region	RG	4	$b_{1,3}>b_{1,5}>b_{1,2}>b_{1,4}>o$	—in Ontario is higher than that in the Prairies
2. Area of residence	AR	3	$b_{2,1}>b_{2,2}>b_{2,3}>o$	—in metropolitan areas is higher than that in small urban areas
				—in the case of head:
3. Education of head	EH	6	$b_{3,7}>b_{3,6}>b_{3,5}>b_{3,4}>b_{3,3}>b_{3,2}>o$	—with a university degree is higher than that with finished elementary schooling
4. Immigration status of head	ISH	3	$b_{4,1}>b_{4,2}>b_{4,4}>o$	—born in Canada is higher than that whose immigration status is not ascertained
5. Main occupation of head in conjunction with his education	MOH.EH	27		
(i) Managerial	MOH_2	4	$b_{5,5}>b_{5,4}>b_{5,3}>b_{5,2}>o$	—with a managerial occupation is higher if the head had a university degree than some or finished elementary schooling
(ii) Professional and technical	MOH_3	4	$b_{5,9}>b_{5,8}>b_{5,7}>b_{5,6}>o$	—with a professional or technical occupation is higher if the head had a university degree than some or finished elementary schooling
(iii) Clerical	MOH_4	3	$b_{5,12}>b_{5,11}>_{5,10}>o$	—with a clerical occupation is higher if the head had some or finished university education than some or finished elementary schooling
(iv) Sales	MOH_5	3	$b_{5,15}>b_{5,14}>b_{5,13}>o$	—with a sales occupation—interpretation as in (iii) above
(v) Service and recreation	MOH_6	3	$b_{5,18}>b_{5,17}>b_{5,16}>o$	—with a service and recreation occupation—interpretation as in (iii) above
(vi) Transport and communication	MOH_7	3	$b_{5,21}>b_{5,20}>b_{5,19}>o$	—with a transportation and communication occupation—interpretation as in (iii) above
(vii) Farmers, farm workers, loggers and fishermen	MOH_8	2	$b_{5,23}>b_{5,22}>o$	—with a specified occupation is higher if the head had more than elementary education than some or finished elementary schooling
(viii) Miners, quarrymen, craftsmen, prod. process and rel. workers	MOH_9	3	$b_{5,26}>b_{5,25}>b_{5,24}>o$	—with a specified occupation—interpretation as in (iii) above
(ix) Labourers	MOH_{10}	2	$b_{5,28}>b_{5,27}>o$	—as a labourer—interpretation as in (vii)

[1] See Appendix A for detailed definitions of symbols.

TABLE 3.1. Sets of Hypotheses Pertaining to Family Income and Their Interpretation – Concluded

Characteristic	Symbol[1]	Number of variables within characteristic (excluding reference variable)	Hypothesis pertaining to each characteristic	Interpretation of hypothesis (taking into account the extreme inequalities as shown in the preceding column)
				Family income – Concluded:
6. Wife's participation in labour force in conjunction with her education	WLF+	8		– in the case of a wife:
(i) Full-year	$WLF+_2$	4	$b_{6,5}>b_{6,4}>b_{6,3}>b_{6,2}>o$	– participating full-year is higher if she had a university degree than that with some or finished elementary schooling
(ii) Part-year	$WLF+_3$	4	$b_{6,9}>b_{6,8}>b_{6,7}>b_{6,6}>o$	– participating part-year is higher if she had a university degree than that with some or finished elementary schooling
7. Head's participation in labour force: full-year or part-year	HLFE	1	$b_{7,1}>o$	– and the head's participation in the labour force is positively related
8. Age of head	AHF	2		
(i) Age in actual years	AHF_1	1	$b_{8,1}>o$	– and the head's age show a parabolic relationship
(ii) Square of AHF_1	AHF_2	1	$b_{8,2}<o$	
9. Proportion of earners in a family	NE/FS	1	$b_{9,1}>o$	– and the proportion of earners in a family are positively related
10. Type of family: all families headed by females and males with marital status other than married	NF_2	1	$b_{10,2}<o$	– in the case of specified families is lower than that of normal husband-wife families

[1] See Appendix A for detailed definitions of symbols.

Exogenous socio-demographic characteristics have been ranked in Chapters 4 and 5 on the basis of their individual contributions to the total variance of income as explained by the model.[9] The technique used in this instance is the analysis of variance which required information on the "explained" or "regression" sum of squares by individual characteristic and its corresponding degrees of freedom. The ratio of the total "regression" sum of squares to the total sum of squares (w.r.t. income) defined the proportion of the total explained variance of income. In order to evaluate the individual contributions of exogenous characteristics to the total explained variance, we required their individual regression sums of squares. Thus for the i-th characteristic with j categories, the regression sum of squares would be the total of individual regression sums of squares pertaining to its (j-1) specified categories (i.e., excluding the reference category). Besides using these individual regression sums of squares corresponding to the exogenous characteristics in evaluating their contributions to the total explained variance, these were used in carrying out the statistical tests to study the significance of each of the socio-demographic characteristics (by applying the familiar analysis of variance technique). Such tests seemed essential as the regression equation dealt simply with the statistical significance of variables and not with the question of over-all significance of each of the broad characteristics.

In conclusion, we may draw attention to the fact that the model postulated in equation (3.1), its related sets of hypotheses and other details relating to the ranking of characteristics and tests of significance, all have been discussed at the national level. As regards the regional analysis, the whole methodological framework is the same except that the model would not include variables pertaining to two exogenous characteristics namely, the immigration status of head and the regional location of family.

[9] A similar method to rank the socio-demographic characteristics has been used by Ostry (44, 45).

CHAPTER 4

EMPIRICAL ANALYSIS OF FAMILY INCOME AT THE REGIONAL LEVEL

Some researchers[1] have already dwelt upon the economics of inter-regional differences in the incomes of families or earnings of individuals existing in Canada. The nature and contents of these studies display a fair amount of conceptual variability.[2] But there is one study, done by Podoluk (46) which needs some special attention primarily because the analytical framework of this chapter is very similar to her exposition. Although both of these studies explain income of an economic family as a function of its basic socio-demographic characteristics as listed in Chapter 2, their statistical methodologies are entirely different. For instance, Podoluk's study consists of statistical tables depicting percentage income distributions by characteristics, ranging from two to four at a time. This form of exposition, although extremely useful for a descriptive presentation of data, requires a good degree of *a-priori* knowledge and perception about the behavioural relationships between the given characteristics.[3] The drawback of this study is that it contains the usual deficiencies of the technique of percentage analysis.[4]

Our methodology, on the other hand, permits us to study the effects of any number of socio-demographic characteristics on the family's income. Put simply, use of multiple regression analysis enables one to quantify the effect of each predetermined variable on the endogenous one by keeping the effects of all others fixed or constant. Further, it helps in studying the statistical significance of the independent variables on income. These are some of the advantages of multivariate regression technique over percentage analysis.

Thus the empirical results, based on multivariate regression analysis, are presented in the following order: (i) a discussion about the overall goodness of fit of the regional models; (ii) the comparative rankings of exogenous socio-demographic characteristics for the individual regions; (iii) a quantitative analysis of pertinent characteristics affecting the family's income and (iv) a statistical procedure to test the equality of the effects of predetermined variables or in technical terms, the equality of estimated regression coefficients over the regions. The detailed regression results for the regions are provided in Tables C.1 to C.5 in Appendix C.

The criteria used to evaluate the goodness of fit of the models are (i) significance of R^2's; (ii) the resulting t-statistics of the estimated coefficients and (iii) the economic interpretation and feasibility (in terms of positive or negative signs) of the estimates in relation to their basic sets of hypotheses.

The estimated values of R^2's (measuring the proportion of variation in the dependent variable that has been explained by the independent variables) were 0.3218, 0.2939, 0.2884, 0.2988 and 0.3322 in the case of the Atlantic, Quebec, Ontario, the Prairies and British Columbia respectively. In terms of magnitudes, these appeared quite interesting because the cross-section models based on dummy variables often yield even lower estimates of R^2's. However, it may be mentioned that these estimates would have been higher if we would have included in the model certain other variables such as one's personal abilities, family background, willingness to take risks, etc. — all of them influencing income.[5]

The statistical significance of these R^2's was tested by the following F-statistic defined as:

$$F = \frac{R^2}{(1-R^2)} \cdot \frac{(N-K-1)}{(K)} \quad \ldots\ldots\ldots (4.1)$$

Where R^2 is the multiple correlation coefficient, N, the number of observations and K, the number of exogenous variables in the equation estimated. This expression indicates that for a given value of R^2, the significance of F-statistic depends heavily on the size of the sample and the number of variables used. Thus using the derived estimates of R^2's and K as 49[6] and with varying N (as shown in Table 2.1), we obtained the corresponding estimates of F for the five regions (quoted in their usual order) as 38.72, 31.95, 40.68, 32.61 and 18.76. All of these turned out to be significant at the 5% level, thus demonstrating the goodness of fit of the regional models.

Individual Contributions and Ranking of Socio-demographic Characteristics

As explained in Chapter 3, we have ranked the exogenous socio-demographic characteristics on the basis of their individual contributions to the total

---

[1] Chernick (12), Denton (18), Brewis (9) and Podoluk (46).

[2] Chernick explained the concept of per capita personal income in relation to certain socio-economic characteristics such as industrial activity, growth in population, rural-urban differences, etc. using weighted coefficient of variation as methodology. Denton used an econometric model based on a time series from 1947 - 64, which consisted of 8 main structural equations. Out of these, two explained the earned income and interest, dividends and net rental income. Brewis, on the other hand, simply described the background factors primarily responsible for the presence of income-differences (using 1961 Census data) between and within regions of Canada. His arguments were somewhat similar to that of Hanna (25), who studied the regional incomes of United States. Finally, Podoluk explained the inter-regional differences in the incomes of economic families in the light of their socio-demographic characteristics. She used 1961 Census data.

[3] Fisher (22) termed this as "discounting *a-priori* the effects of independent variates". p. 164.

[4] Zeisel (59) has described in details some major drawbacks of percentage analysis.

[5] It may be noted that an estimate of R^2 is quite sensitive to (i) the errors of measurements of the exogenous variables; (ii) various sampling and non-sampling errors in the endogenous variable; and (iii) the specification of the model. See Cochran (16) for a lucid study of the effects of errors of measurement on multiple correlation.

[6] Although the regional model contained 55 independent variables (excluding the constant term), the exclusion of 6 reference categories from this resulted in 49 variables in the final version for estimation purposes.

explained variance of income. Table 4.1 describes such contributions to the explained variances for the individual regions. As regards their estimation, the contribution of the i-th characteristic (i = 1, 2, ... 8) for the k-th region (k = 1, 2, ... 5) has been the ratio of the regression sum of squares pertaining to the i-th characteristic for the k-th region to the total sum of squares (w.r.t. income) for the k-th region.

The table revealed that area of residence accounted for almost 1/7th and 1/6th of the total explained variance of income (hereafter referred to as TEVI) in the Atlantic and the Prairie provinces respectively. Similarly, the education of the head of family explained nearly 1/5th, 1/7th, 1/5th, 1/10th and 1/6th of the TEVI for the Atlantic, Quebec, Ontario, the Prairies and British Columbia respectively. But the most interesting feature of this table was the revelation about the contribution of MOH.EH; i.e., the head's occupation in conjunction with his education, accounting for about 50% of TEVI in the case of Ontario compared to 47% and 29% for Quebec and the Atlantic region. Similarly the contribution corresponding to the proportion of earners varied over the regions; for example, it had the highest share in Quebec and the Atlantic region; i.e., 16% and 14% of their respective TEVI's.

TABLE 4.1. Distribution of Explained Variance by Exogenous Socio-demographic Characteristics by Regions

Characteristic	Atlantic	Quebec	Ontario	Prairies	British Columbia
Area of residence (AR)	0.0472	0.0239	0.0171	0.0463	0.0097
Education of head (EH)	0.0690	0.0415	0.0582	0.0274	0.0523
Main occupation of head in conjunction with his education (MOH.EH)	0.1242	0.1372	0.1492	0.1489	0.1531
Wife's participation in labour force in conjunction with her education (WLF+)	0.0156	0.0053	0.0072	0.0139	0.0280
Head's participation in labour force (HLF)	0.0010	0.0108	[1]	[1]	[1]
Age of head (AHF)	0.0107	0.0248	0.0209	0.0169	0.0418
Proportion of earners (NE/FS)	0.0447	0.0463	0.0272	0.0255	0.0332
Type of family (NF)	0.0094	0.0041	0.0086	0.0199	0.0141
Total explained variance	**0.3218**	**0.2939**	**0.2884**	**0.2988**	**0.3322**

[1] Figures not appropriate as the regression coefficients of HLFE showed unrealistic and economically implausible results in the case of Ontario, the Prairies and British Columbia, their corresponding regression sums of squares and their related ratios were not estimated. See Tables C.3 to C.5.

Taking into account certain other variables such as the wife's participation in labour force, nature of work of head and his age, it must be observed that the relative contributions pertaining to the wife's participation in labour force in conjunction with her education and age of head were highest in British Columbia compared to all other regions. As regards the performance of the characteristic HLF, we don't think we could argue that this did not influence income of a family living in Ontario or British Columbia but as the empirical analysis stood,[7] its odd situation could be ascribed to its related problems of grouping as elaborated in Chapter 2.

It would be more appropriate to rank these characteristics according to their individual contributions to the TEVI in order to demonstrate their importance in the different regions. Data from Table 4.1 has been re-arranged and are presented in Table 4.2 in ranked order.

[7] See footnote 1, Table 4.1.

TABLE 4.2. Ranking of the Socio-demographic Characteristics in Descending Order of Their Individual Contribution to Total Explained Variance of Income

Atlantic	Quebec	Ontario	Prairies	British Columbia
MOH.EH	MOH.EH	MOH.EH	MOH.EH	MOH.EH
EH	NE/FS	EH	AR	EH
AR	EH	NE/FS	EH	AHF
NE/FS	AHF	AHF	NE/FS	NE/FS
WLF+	AR	AR	NF	WLF+
AHF	HLF	NF	AHF	NF
NF	WLF+	WLF+	WLF+	AR
HLF	NF	HLF[1]	HLF[1]	HLF[1]

[1] Placed at the end because estimates could not be interpreted.

Source: Table 4.1.

The most noteworthy feature revealed by this table is that none of the regions had an identical ranking of all these socio-demographic characteristics. It has also demonstrated that out of all exogenous characteristics, the head's main occupation in conjunction with his education was a predominant factor in determining the family's income over all the regions. Following this, the education of head alone seemed economically vital in the case of Atlantic, Ontario and British Columbia compared to the proportion of earners in Quebec.[8] As regards the area of residence, it proved highly influential in respect to the income of a family living in the Prairies compared to its lowest ranking in the case of British Columbia.[9] On the other hand, the age of the head of family was most important in British Columbia compared to all other regions. The head's participation in labour force turned out to be of greater significance in relation to the income of a family residing in Quebec. Similar comparisons pertaining to other socio-demographic characteristics influencing family income at the regional level can be drawn from Table 4.2.

The statistical significance of the qualitative socio-demographic characteristics with more than two categories has been tested by the well known analysis of variance technique.[10] The resulting F-ratio pertaining to the i-th characteristic was derived by dividing the mean sum of squares pertaining to the i-th characteristic by the mean error sum of squares (the mean sum of squares was obtained by dividing the sum of squares by the corresponding degrees of freedom). Such ratios for the selected characteristics are given below:

TABLE 4.3. The Estimated F-values for Selected Characteristics

Characteristic[1]	Atlantic	Quebec	Ontario	Prairies	British Columbia
AR	90.94	41.48	38.35	77.14	8.52
EH	66.45	36.02	65.25	22.82	22.96
MOH.EH	26.59	26.46	37.17	27.55	14.94
WLF+	11.27	3.44	6.06	8.68	9.22

[1] See definitions of characteristics in Table 4.1.

All these F-ratios proved significant at the 5% level (taking into account the corresponding degrees of freedom) indicating the overall significance of AR, EH, MOH.EH and WLF+ to the family's income in all regions.

At a more detailed level, we examined the individual contributions of the 49 variables specified in the regional model (excluding the reference variables or categories) to the total explained variance of income.[11] This was done by taking into account the ratios of the corresponding regression sums of squares to the total sum of squares (w.r.t. income). The study showed that the professional and technical group in conjunction with university education ranked first in the Atlantic, Quebec, Ontario and British Columbia. Whereas in the Prairies, metropolitan area topped the list of exogenous variables followed by the above mentioned occupation-education variable. Similarly, out of the educational groups, the university graduated head appeared to be making a significant contribution to the family's income in the case of Ontario and British Columbia. Another result of this process of ranking variables was the emergence of the variable defining the proportion of earners in a family in the fifth place (out of 49) in Ontario, the Prairies and British Columbia.

[8] According to the 1961 Census, the average size of family is larger in Quebec than in other regions. Moreover, the average number of children in the older age group, i.e., 16-25 years, staying with parents is also highest in this region. So our inference would be true in the light of the fact that when these children participate in the labour force, they also become income earners, thus, increasing the proportion of earners in an economic family. This in turn seems to contribute significantly to the family's income. For further details, refer to Table 88, *Households and Families*, Census of Canada, 1961, Vol. 2, Part 1, Dominion Bureau of Statistics, Ottawa.

[9] This could be due to the fact that a greater proportion of families in British Columbia lived in metropolitan and other urban areas. So their incomes would be less affected by the nature of their area of residence compared to other important characteristics that show more variation (e.g., age of head).

[10] The statistical significance of quantitative variables, namely AHF and NE/FS is established by the t-values of their coefficients. As regards the other two characteristics, i.e., HLF and NF, each of these having only two categories, the analysis of variance test need not be applied to them. The t-values of the estimated coefficients are the appropriate criteria to judge the significance of these characteristics.

[11] The estimates of individual contributions are not supplied in the text but can be made available on request (see footnote 3 of Chapter 3).

Quantification of the Effects of Variables Pertaining to Socio-demographic Characteristics

Almost all of the hypotheses, stated in Table 3.1, both in terms of signs and numeric order of regression coefficients, were fully or partly satisfied for all the regions.[12] The only characteristic which did not meet our expectations was HLF,[13] where only in Quebec and the Atlantic region, we could accept the stated hypothesis. Coefficients for variables EH_2, EH_3, EH_4, EH_6, MOH_{61}, MOH_{81}, MOH_{82}, MOH_{101}, MOH_{102}, $WLF+_{21}$ and HLFE– were all non-significant at the 5% level in all regions.

The most important of the socio-demographic characteristics was the head's main occupation in conjunction with his education. As regards the occupation, the column vectors of the means ($\bar{X}_{i,j}$) of Tables C.1 to C.5 identified three major occupational groups, namely (i) the managerial, (ii) professional and technical and (iii) miners, craftsmen, production process and related workers. Combined these accounted for 44.8%, 50.7%, 56.8%, 42.3% and 52.6% of the total occupational distribution of heads of families in the case of the Atlantic, Quebec, Ontario, the Prairies and British Columbia respectively. This distribution further revealed that almost one out of every five family heads was either a farmer or a farm worker in the Prairies.[14]

The corresponding regression coefficients of the variables pertaining to the above listed occupational groups taken in conjunction with the head's education revealed interesting features. As expected, the incomes within these occupations increased with the level of education. For example, an Ontario family with a head with finished elementary education and engaged in managerial occupation would show an income difference of $5,282 compared to a family whose head had no education and did not work.[15] Similarly, an Ontario family with a head in the same occupational group but with a university degree would have an income $9,799 higher than the reference group mentioned earlier. In other words, two levels of education (finished elementary education and a university degree) of the head within the same occupational group would result in a difference of $4,517 in the incomes of Ontario families. The corresponding differences for the Atlantic, Quebec, the Prairies and British Columbia were $3,626, $4,949, $6,022 and $2,726 respectively. On the other hand, the professional and technical group yielded differences in the incomes of families accruing to the various educational levels of their heads (i.e., finished elementary schooling versus finished university education). For example, these amounted to $4,589, $3,859 and $3,791 in the case of Quebec, Ontario and the Prairies. Lastly, considering the occupational group of miners, craftsmen, production process and related workers, the above mentioned levels of education of the heads of families resulted in income differences of $1,420, $1,495, $623, $1,042 and $292 in the case of the Atlantic, Quebec, Ontario, the Prairies and British Columbia respectively.

A comparative study of income differences, as illustrated above, indicated that varying levels of education would result in higher differences in the incomes of families whose heads were engaged in jobs requiring greater skills and other technical know-how. This inference conformed to Mincer's (37) findings that occupations requiring higher education received higher incomes, which in turn, varied not simply by an additive constant but by a multiplicative factor. This statement can be tested by considering the professional and technical group, for example, in the Prairies. The estimated regression coefficients of MOH_{31}, MOH_{32} and MOH_{34} were $1,648, $2,600 and $5,439. The income difference resulting from finished elementary to finished high school education was $952 compared to a difference of $2,839 resulting from finished high school education to the graduation from university; the latter difference being almost three times the former. Similar other progressions of income differentials pertaining to different levels of education could be studied for all other regions.

Next, let us consider the inter-regional income variability as depicted by the regression coefficients pertaining to MOH.EH. We measured such variability by calculating the coefficient of variation defined as a ratio of the standard deviation to the mean of the coefficients.[16] As an illustration, we selected four occupational groups, namely the managerial, professional and technical, sales and the labourers. The first two represent jobs requiring special skills whereas the latter two can be described as unskilled jobs. The coefficients of variation with respect to MOH_{21}, MOH_{24}, MOH_{31} and MOH_{34} were 0.498, 0.274, 0.749 and 0.105. Similarly, the coefficients of variation pertaining to MOH_{51}, MOH_{53}, MOH_{101} and MOH_{102} were 0.388, 0.139, 0.623 and 0.384. A comparison of these sets of coefficients of variation led us to conclude, first, that the inter-regional income differentials (another way of interpreting the estimated regression coefficients as these denoted the differences with respect to their reference variables) were low in the case of skilled occupational groups; they decreased with increases in the head's level of education. Secondly, the variability between the estimated income

[12] See Chapter 3 for a definition of fully and partly satisfied hypotheses. A table summarizing these results is given in Chapter 6, Table 6.1.

[13] See Appendix A for the definitions of variables.

[14] Although our definition of MOH_8 includes fishermen and loggers along with farmers and farm workers, its representation in the Prairies is accounted for overwhelmingly by farmers and farm workers. Loggers and fishermen in this case accounted for only 0.11%.

[15] In the following paragraphs, income differences always refer to (except few instances as specified) the income of a given category compared with the reference category as defined in Appendix A.

[16] See Table C.1 to C.5 for the regression coefficients.

differentials accentuated with a decline in skill requirements of job; i.e. in the case of unskilled occupational groups.[17]

We have so far discussed the inter-regional variability between the estimated coefficients pertaining to the categories of certain selected occupation groups. But another feature which equally interested us was to determine the degree of inter-regional variability attributable to the main occupation groups. In this context, we determined the weighted coefficient of variation[18] for the occupations namely the managerial, professional and technical, sales and the labourers. Their corresponding coefficients, 0.259, 0.167, 0.301 and 0.550 suggested that the overall variation between incomes of families headed by the engineers, teachers, professors, lawyers, doctors and other professionals, over the regions is smaller than the inter-regional variation in family incomes of unskilled workers. This could be due to the fact that these professional and technical people are somewhat more homogeneous over all regions.

The characteristic which ranked second in the case of the Atlantic, Ontario and British Columbia and third in the Prairies and Quebec was the education of the head of family. A comparative study of estimated regression coefficients pertaining to EH_7 revealed the highest gain in the income of an Ontario family with a head possessing a university degree compared to a family whose head had no education. These amounted to $3,722 compared to the corresponding income gains of $3,260, $2,697, $2,628 and $1,529 in the case of British Columbia, Quebec, the Atlantic and the Prairies. Similarly, the differences of the estimates corresponding to EH_7 and EH_5 over the regions, which were significant at 5% level, amounted to $2,822 and $2,608 in Ontario and British Columbia. In other words, the level of education of the heads, i.e., whether they had university degrees or not, considerably affected incomes of the families living in these regions.

The coefficients of variation corresponding to EH_3, EH_5 and EH_7 were 0.488, 0.298 and 0.266 respectively. These estimates, in turn, suggested that inter-regional dispersion of income was maximum in the case of families with heads with completed elementary education and minimum in the case of heads possessing university degrees. In other words, the regional income differentials decreased as the level of education increased.

We may emphasize that much inter-regional differences in the incomes of families primarily arose from the degree of skewness in the distributions of families by levels of education of their heads. For instance, one out of every two heads in the case of the Atlantic families had some or finished elementary education; the highest proportion among all the regions. On the other hand, 50% of the heads of Ontario families had some or finished high school education compared to 38%, 39%, 47% and 54% in the Atlantic, Quebec, the Prairies and British Columbia. The lowest proportion of families with university graduated heads were found in the Atlantic region.

In order to demonstrate the relationship between the distribution of families by levels of education of heads in a region and its corresponding income level, we evaluated the rank correlation coefficients. Some of these turned out to be as follows:[19]

(i) between no education and family income up to $2,999 = 0.90

(ii) between university degree and family income of $10,000 and over = 0.90

The significance of these correlations at the 5% level demonstrated how closely the regional income levels and the educational patterns of their populations are related.

The proportion of earners ranked third in Ontario and second in Quebec. Comparison of Tables C.1 to C.5 showed that Quebec had the highest coefficient, implying an increase of $3,351 in the family's income in relation to a unit increase in the proportion of earners. This highest estimate was followed by other incremental incomes amounting to $2,592, $2,243, $1,997 and $1,933 in the case of Ontario, the Atlantic, the Prairies and British Columbia respectively. These coefficients, in turn, revealed an inter-regional variability of 21.4% between the incomes of families with respect to their proportions of earners.

Age of the head of family, which ranked third in British Columbia, appeared in the fourth place in Ontario and Quebec. This could partly be attributed to the differences in the age-compositions of the populations. For example, British Columbia had the lowest proportion of families with heads in the age group 55 - 64 years and highest proportion of families in the age group 65 years and over. Similarly, the significance of age in terms of the family's income could be influenced by the economic activities of the regions; e.g., the predominance of manufacturing activity would have a different impact on the age structure of the working population (and on family incomes) than that of either agriculture or service industries.

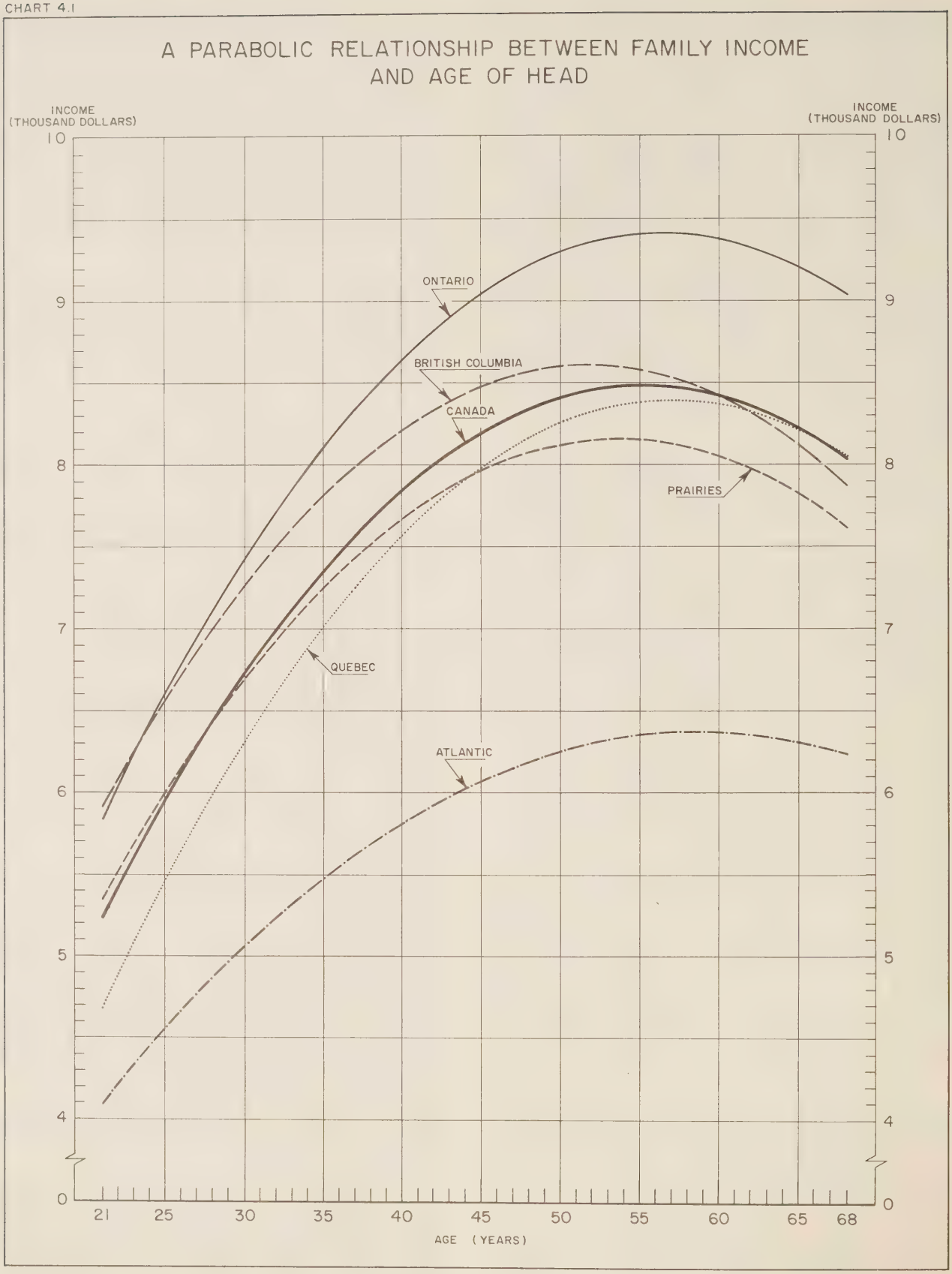
---

[17] This conclusion is similar to the one derived by Podoluk (46), p. 168. In another study, Becker (3) placed strong emphasis on vocational training as a factor determining the dispersion and skewness of earnings.

[18] A weighted coefficient of variation for the i-th characteristic, as used in this text, is defined as $(N_{i1}V_{i1} + N_{i2}V_{i2} + \ldots + N_{ij}V_{ij})/N_{i1} + N_{i2} + \ldots + N_{ij}$, where V_{ij} is the coefficient of variation for the j-th variable pertaining to the i-th characteristic and N_{ij} is the relative proportion of the j-th variable of the i-th characteristic over all the regions. In other words, N_{ij}'s were taken from the national model.

[19] Families were grouped into four broad income groups namely (a) up to $2,999, (b) $3,000 - 6,499, (c) $6,500 - 9,999 and (d) $10,000 and over in the five regions. The rank correlations were derived on the basis of this table and the column vectors of means ($\bar{X}_{i,j}$) of Tables C.1 to C.5. Not all of these correlations were statistically significant.

CHART 4.1


A PARABOLIC RELATIONSHIP BETWEEN FAMILY INCOME AND AGE OF HEAD
INCOME (THOUSAND DOLLARS)
INCOME (THOUSAND DOLLARS)
10
9
8
7
6
5
4
0
ONTARIO
BRITISH COLUMBIA
CANADA
PRAIRIES
QUEBEC
ATLANTIC
21
25
30
35
40
45
50
55
60
65
68
AGE (YEARS)

As stated in Chapter 2, we included the squared term of the age variable as an exogenous variable in the model with an objective to determine the optimal age of the head of family when its income would achieve its maximum. The corresponding income-age profiles for the individual regions as derived from the models (keeping the effects of all other variables constant) are shown in Chart 4.1. The analysis revealed that an Ontario family would have the maximum income by the time its head had reached the age of 56.6 years. Similarly, families in the case of the Atlantic, Quebec, the Prairies and British Columbia achieved their income peaks by the time the heads were 58.8, 57.0, 53.8 and 51.9 years old respectively. Like other studies,[20] Chart 4.1 confirmed the widely accepted notion that incomes increase up to a certain age beyond which they begin to decrease.

A comparison of inter-regional regression coefficients pertaining to the age of head indicated the highest increase of $327 attributable to an additional year in the head's age (alternatively, acquisition of experience and skills, etc.) in Quebec compared to $322, $294, $283 and $188 in Ontario, British Columbia, the Prairies and the Atlantic region. These resulted in a coefficient of variation of 0.177, which was comparatively smaller than the one corresponding to the proportion of earners. In other words, the inter-regional income dispersion with respect to the age of head of family was smaller than that due to the proportion of earners.

The wife's participation in labour force ranked fourth in the case of the Atlantic region and British Columbia and seventh in Quebec, Ontario and the Prairies. The mean vectors of Tables C.1 to C.5 revealed that Ontario and British Columbia had the highest proportions of wives participating both full-year and part-year, in the labour force; i.e., 37% and 35% respectively. It was also interesting to observe that 17.5% of the wives participated on full-year basis in the Atlantic region compared to only 9.1% in Quebec.

The positive and significant regression coefficients indicated that wives working all year accounted for income differentials in all the regions. Especially in British Columbia and the Prairies, families with wives with some university education and working full-year made income gains of $2,779 and $2,331 compared to the families with wives who had no education and who did not participate in the labour force. Income differentials of $1,249, $278 and $893 were found in the case of the Atlantic region, Quebec and Ontario respectively. Taking into account the differences in the levels of education of wives, we found that wives with a university degree working full-year accounted for income differentials of $1,891 and $1,010 compared to those with finished high school education in British Columbia and the Prairies. On the other hand, wives with a university degree and participating on a part-year basis made the highest contribution ($1,940) to the family's income in the Atlantic region.

On the whole, the inter-regional dispersion of income associated with the wife's full-year participation showed unexpected result. It was 0.148 in the case of wives with finished high school education and 0.633 for university graduated wives. This implied that income differentials accruing to the wife's participation in labour force increased with an increase in her level of education. This is in direct contrast to what we found in the case of the head's education as related to the inter-regional dispersion of family incomes. The reasons for these phenomena have not been explored in this study.

The area of residence ranked second as an explanation of income variability in the Prairies and fifth in the case of Quebec and Ontario. The mean vectors of Tables C.1 to C.5 revealed that the Atlantic region had the highest proportion of families living in the rural areas compared to other regions of Canada. On the other side of the coin, Ontario had the highest proportion of families residing in metropolitan areas.

Such distribution of families by area of residence in itself would have indicated the existence of the differences in their incomes. Let us now examine to what extent they were different. The estimated coefficients revealed that the differences in family incomes due to the place of residence, i.e., between metropolitan and rural areas, amounted to $1,802, $1,447, $1,388, $1,155 and $1,053 in the Prairies, the Atlantic region, Ontario, Quebec and British Columbia respectively. In other words, the income difference with respect to the above mentioned areas of residence was highest in the Prairies and lowest in British Columbia.

The inter-regional dispersion of income in relation to area of residence was smallest in the case of metropolitan areas and increased as we moved to other cities or small urban areas. In empirical terms, the coefficient of variation with respect to metropolitan areas was 0.190 compared to 0.276 in the case of other small urban areas. This inference[21] appeared to be consistent with Podoluk's (46) statement that "much less regional variation existed in urban incomes than in rural incomes; the main contrast was between the Atlantic region and the remaining provinces", p. 160.

Lastly the empirical analysis revealed that the type of family would also account for income differentials. For example, the difference in the incomes of husband-wife family and other type of family was highest in the Prairies, i.e., $1,455. Similar other income differentials were $1,436, $1,344, $858 and $782 in the case of British Columbia, Ontario, the Atlantic region and Quebec. The differences of the

[20] Morgan, et. al. (40), Mincer (39), Lydall (34) and David and Miller (17). Especially see Lydall, pp. 113-25 where the author listed several characteristics changing over time, which influence one's earnings. For example, "dispersion of ability . . . . tends to increase with age, both absolutely and relatively. . ." p. 114.

[21] See Chernick (12), p. 45.

estimates pertaining to British Columbia and Quebec and Ontario and Quebec, for example, were significant at the 5% level. Such income differences could be attributed to the differences in the composition of families and their other socio-demographic characteristics which have not been taken into account here.

Testing the Equality of Regression Coefficients

Let us now proceed to test whether the specified exogenous variables had equal effects on the family's total income in all the regions. The statistical test used in this context is briefly outlined below.

Defining the original model, as postulated in equation (3.1) for the i-th region in matrix notations as:

$$Y_i = X_i B_i + U_i \qquad (i = 1,2,..5) \qquad ...(4.2)$$

where X_i is a matrix of order $(N_i \times K)$ and Y_i, B_i and U_i are column vectors of order $(N_i \times 1)$, $(K \times 1)$ and $(N_i \times 1)$ respectively. N_i denotes the sample size for the i-th region (as shown in Table 2.1) and K, the number of exogenous variables (including the constant term) in the model.

Then our null hypothesis states that all the B_i's are equal; i.e.:

$$B_1 = B_2 = B_3 = B_4 = B_5 \qquad ...(4.3)$$

where the suffixes 1,2, . . .5 represent the regions namely, the Atlantic, Quebec, Ontario, the Prairies and British Columbia. In conventional terminology, such hypothesis is tested against either (a) some *a-priori* knowledge about these unknown coefficients or (b) the derived vector of estimated coefficients from pooled sample data. In the absence of any *a-priori* information about such unknown parameters, we opted to follow (b) and obtained a vector of estimated coefficients, $\hat{B}_p$, the suffix p indicates pooling of all the regional samples. Thus the final version of the null hypothesis takes the form of:

$$B_1 = B_2 = B_3 = B_4 = B_5 = B_p \text{ (say)} \qquad ...(4.4)$$

In technical terms, $\hat{B}_p$ represents the common regression plane. In view of the fact that statistical details to test such hypothesis has been discussed at length in various studies,[22] it would be sufficient for our purposes to simply delineate the required F-statistic as described by Huang (28).

$$F = \frac{(Q_1' - Q_2')/4K}{Q_2'/(N-5K)} \qquad ...(4.5)$$

where

$$Q_1' = (Y_p - X_p \hat{B}_p)'(Y_p - X_p \hat{B}_p)$$

$$Q_2' = (Y_1 - X_1 \hat{B}_1)'(Y_1 - X_1 \hat{B}_1) + (Y_2 - X_2 \hat{B}_2)'(Y_2 - X_2 \hat{B}_2) + ... + (Y_5 - X_5 \hat{B}_5)'(Y_5 - X_5 \hat{B}_5) \qquad ...(4.6)$$

In simple words, Q_1' measured the residual sum of squares derived from the pooled sample and Q_2' denoted the sum of residual sums of squares derived from the samples of individual regions. As usual, N and K denote the pooled sample size and the number of exogenous variables including the constant term.

Substituting the values of N and K as 18,143 and 50 (excluding the number of reference variables in the model) and Q_1' and Q_2', equation (4.5) yielded an estimate of F-statistic as 2.27. This in turn was greater than the tabulated value of F-statistic at (4K, N-5K) degrees of freedom and 5% level of significance. Hence we rejected the null hypothesis.

In sum, Table 4.2 and the above test have confirmed that the given socio-demographic characteristics and their related variables would have different effects on the incomes of families living in different regions. It was rather interesting to observe from this multivariate study that in spite of the overall significance of these variables on the family's income, none of them had identical quantitative effects (in terms of regression coefficients) in any two regions.

Thus, in conclusion, we may add that all statistical criteria established the goodness-of-fit of the regional models. The empirical analysis based on these, in fact, has enabled us to explore certain hypotheses which could not have been tested by simple two-way or three-way cross tabulations.

---

[22] Goldberger (23) and Chow (15).

CHAPTER 5

EMPIRICAL STUDY OF FAMILY INCOME AT THE NATIONAL LEVEL

The income model at the national level is basically a replica of the regional model except that it contains two additional exogenous socio-demographic characteristics. These are (i) the regional location of the family and (ii) the immigration status of its head. The background and usefulness of these characteristics in the model have been explained in Chapter 2. The categories of these characteristics were fed into the model in terms of dummy variables.

The national model was tested for goodness of fit and accuracy by the same criteria as applied in the case of the regional models.[1] Its R^2 turned out to be 0.2992, which, on the basis of the F-statistic described in equation (4.1), turned out to be significant at the 5% level.[2]

The individual contributions of the exogenous characteristics to the total explained variance of income (TEVI) were quantified by the analysis of variance technique as discussed in the previous chapter. The results revealed that 47% of TEVI accrued to the cross-classified characteristic MOH.EH. On the other hand, nearly 1/9th of TEVI was attributed to the proportion of earners in a family compared to nearly 1/11th ascribing to the area of residence. Besides these contributions, the age of head and the regional location of family accounted for nearly 7% and 4% of TEVI respectively.

Thus a rearrangement of these characteristics on the basis of their individual contributions to the TEVI (in descending order of their magnitudes) gave the following order of characteristics: (i) MOH.EH, (ii) EH, (iii) NE/FS, (iv) AR, (v) AHF, (vi) RG, (vii) WLF+, (viii) NF, (ix) HLF and (x) ISH.

Such ranking of characteristics has, therefore, demonstrated that the most pertinent characteristic affecting a family's income in Canada is the head's main occupation in conjunction with his education. The head's education, as a separate variable, also played a significant role in determining family incomes. It was interesting to note that area of residence influenced the family's income more strongly than its regional location. Similarly, the proportion of earners in a family and the age of head appeared to have greater influence on a family's income than other characteristics such as RG, WLF+, HLF and NF. On the other hand, it was rather baffling to find the immigration status of the family head at the bottom of the list.[3]

The overall significance of the exogenous characteristics, namely, RG, AR, EH, MOH.EH and WLF+ was also tested by evaluating the corresponding F-ratios (using analysis of variance technique described in Chapter 4). These turned out to be 76.11, 235.92, 196.26, 135.53 and 26.75 respectively. Needless to say, all of these proved significant at the 5% level with the corresponding degrees of freedom.

Taking into account the individual contributions of variables (excluding the reference variables) to the TEVI, we found that MOH_{34} ranked first out of 56 variables of the estimated income model. This implied that a head having a university degree and belonging to a professional and technical occupational group would have a greater impact on family income than heads in all other categories of the occupational-educational classification. Other revealing features emanating from a comparative study of individual attributes were (i) the proportion of earners in a family ranked third and (ii) the age of its head in seventh place – both of these were quantitative variables in the model. As regards the area of residence and regional location of the family, residence in a metropolitan area ranked sixth and living in Ontario in the eighth position. In sum, the major attributes influencing income of a Canadian family were (i) the head's university education and belonging into a professional and technical occupation, (ii) the proportion of earners in a family, (iii) residence in a metropolitan area, (iv) living in Ontario and (v) age of the head of the family.

Evaluating the national model in terms of meaningful interpretations of the regression coefficients and their related hypotheses, we found that almost all of the stated hypotheses were fully-satisfied.[4] However, the variables – EH_2, EH_3, EH_4, EH_6, MOH_{61}, MOH_{81}, MOH_{82}, MOH_{102}, $WLF+_{21}$, $WLF+_{34}$ and HLFE were statistically non-significant at the 5% level.[5]

The mean vector ($\bar{X}_{i,j}$) of Table C.6 revealed that 1/10th of the heads of Canadian families were in professional and technical jobs compared to 30% belonging to a group of miners, craftsmen, production process and related workers. The estimated regression coefficient pertaining to MOH_{34} showed that a Canadian family whose head had a university degree and was in professional and technical occupation would have $5,289 more income than a family whose head had no education and did not work. Similarly, a family with a head with a university degree in a managerial occupation would have an income higher by $8,155 than the reference group mentioned above. The difference between these two income increments, i.e., $2,866 was the income difference due to occupation for families

[1] See Chapter 4 for their details.
[2] The values of N and K in this case were 18,143 and 56 (excluding eight reference variables) respectively.
[3] Its individual contribution to TEVI was almost negligible. In the absence of any logical explanation this situation must be accepted as an empirical oddity.
[4] See footnote 12 of Chapter 4.
[5] See footnote 13 of Chapter 4.

whose heads had university degrees. Families with heads in managerial occupations had higher incomes by that amount compared to families with heads in the professional and technical occupations – this was statistically significant at the 5% level.

The regression coefficients with respect to EH showed that a Canadian family with a head with a university degree would have higher income than the one with head with finished elementary and high school education. For instance, the income difference accruing to the change from finished elementary education to a university degree was $2,462 compared to $1,984 resulting from a change from finished high school education to a university degree. These income differences arising from the changes in the levels of education, in turn, proved significant at the 5% level.

The empirical analysis also revealed that the income of a Canadian family would reach its peak by the time the head was 55 years of age. Up to that age, each additional year in the head's age (i.e., skill and experience gained over time) would bring an increase of $307 in the family's income. On the other hand, a unit increase in the proportion of earners in a family would augment its income by $2,624.

Taking into account the area of residence and regional location of a family, we found that about 61% of the Canadian families lived in metropolitan areas whereas only 1/5th lived in rural areas. Moreover, a family residing in metropolitan area had higher income than the one in the rural area. The corresponding income differential was nearly $1,412. Similarly, the difference in family income between metropolitan and small urban areas was $971.

Nearly 37% of the total Canadian families lived in Ontario compared to 27.6% in Quebec and 10% in British Columbia. In terms of income differentials, the income of an Ontario family was by $1,312 higher than that of a family in the Atlantic region. Similarly, in British Columbia, Quebec and the Prairies, family incomes were respectively higher by $870, $796 and $474 than those in the Atlantic region.

As regards the contribution of wives, the analysis revealed that a Canadian family with a wife who had some university education and full-year participation in the labour force had the highest income differential – $1,409 compared to a family with a wife with no education and who did not work. The mean vector of Table C.6 showed that only 21.9% of all husband-wife families had wives participating full-year in the labour force. Out of this, 64% of the wives had only completed high school education compared to about 4% with a university degree. The contribution made by a wife with a completed high school education was only a quarter of that made by a wife with some university education. On the other hand, the income increment in the case of a wife with a university degree was comparatively lower than that of a wife with some university education. This was contrary to our *a-priori* hypothesis pertaining to this characteristic.

The difference in the income of a husband-wife family compared to all other types of families turned out to be $1,133, which in turn, was significant at the 5% level. This implied that in Canada, the type or structure of family would influence the magnitude of its total income. Such income differences between a husband-wife family and all other types of families could be attributed to (a) stronger motivation and (b) economic potential (a function of certain socio-demographic characteristics not included in the model) of the husband-wife families.

In conclusion, mention must be made that the estimates derived from the national model would more-or-less represent the averages of the regional estimates presented in Chapter 4. It must, however, be borne in mind that direct comparison of estimates pertaining to variables common in both the regional and the national models cannot be made because the latter model specifies the partial effects of one exogenous characteristic (i.e., the immigration status of the head of family) explicitly. But as regards the significance of socio-demographic characteristics on the family's income, both the national and the regional models in broad terms yielded consistent results.

CHAPTER 6

CONCLUSIONS

We have endeavoured to present in this paper a simple econometric study of income of an economic family both at the regional and the national levels. The primary objective of this study was to rank socio-demographic characteristics affecting income in an order of importance and also quantify their differential effects on the family's income. The income variable under study was the "total family income" consisting of various components of cash income, (i.e. earned and unearned income).

To achieve this objective, we used the 1967 income data collected in April 1968 for the Consumer Finance Research Staff. From these data we extracted the following socio-demographic characteristics: (i) regional location of family, (ii) area of residence, (iii) sex of head, (iv) education of head, (v) main occupation of head, (vi) wife's participation in labour force, (vii) number of weeks worked by head, (viii) age of head (in grouped form), (ix) immigration status of head, (x) number of children (by age groups, i.e. below 6 years, 6 - 15 years and 16 - 25 years old) and (xi) the proportion of earners in a family. In view of the fact that (v), (vi), (vii), (viii) and (x) created certain technical problems, their specification in the model needed some special treatment. This resulted in amended specification of (a) main occupation of head in conjunction with his education, (b) wife's participation in labour force in conjunction with her education (applicable only in a husband-wife family), (c) age of head in actual years and inclusion of its squared term in order to present a more realistic income-age profile, (d) head's labour force participation in terms of broad groups, i.e. whether he participated or not and (e) the type of family, i.e. husband-wife family versus single parent and all other types of families.

Thus the final version of the single equation income model (at the regional level) contained eight exogenous socio-demographic characteristics namely, (i) area of residence, (ii) education of head, (iii) main occupation of head in conjunction with his education, (iv) wife's participation in labour force in conjunction with her education, (v) head's labour force participation, (vi) age of head, (vii) the proportion of earners in a family and (viii) the type of family. The two additional characteristics at the national level were (ix) the regional location of the family and (x) immigration status of its head. Age of head and proportion of earners were fed into the models in quantitative terms whereas all other qualitative characteristics were transformed into dummy variables. Thus including the constant term, we had 56 variables in the regional model compared to 65 at the national level. In order to use the method of simple least squared estimation, we further excluded one variable from each of the sets of variables pertaining to the given socio-demographic characteristics. Such excluded variables were treated as reference variables (or categories). This way the regression coefficients of variables were interpreted in terms of differences with respect to their reference variables.

The goodness-of-fit of an individual model was judged (i) on the basis of the percentage proportion of the total explained variance of income derived from the model and its statistical significance and (ii) in terms of the number of stated hypotheses pertaining to the exogenous socio-demographic characteristics.

As regards the percentage proportions of the total explained variances of incomes, the estimates were 32.18%, 29.39%, 28.84%, 29.88%, 33.22% and 29.92% in the case of the Atlantic, Quebec, Ontario, the Prairies, British Columbia and all Canada respectively. All of these estimates turned out to be significant at the 5% level.

The following table summarizes the distribution of tested hypotheses under three classifications:

TABLE 6.1. Empirical Results in Respect to Hypotheses Stated in Table 3.1

Classification	Atlantic	Quebec	Ontario	Prairies	British Columbia	Canada
Fully-satisfied[1]	10	11	8	11	7	15
Partly-satisfied[2]	3	2	6	4	7	3
Not satisfied	4	4	3	2	3	1
Total[3]	**17**	**17**	**17**	**17**	**17**	**19**

[1] Hypothesis pertaining to the i-th characteristic is fully-satisfied if the numeric order of the estimated regression coefficients followed our *a-priori* specification of b's corresponding to the i-th characteristic.

[2] Hypothesis pertaining to the k-th characteristic is partly-satisfied if half or more than half of the estimated regression coefficients followed our *a-priori* specification of b's corresponding to the k-th characteristic.

[3] Difference in the regional and the Canada totals is due to the additional characteristics, namely the region and the immigration status of the head of family, in the national model.

Examining further these sets of hypotheses, we found that the hypotheses pertaining to (i) the proportion of earners in a family, (ii) age of head and (iii) type of family, were fully satisfied in all the regions as well as at the national level. On the other side of the coin, the hypothesis pertaining to the wife's part-year participation in labour force in conjunction with her education was not satisfied in the case of either region or Canada.

The socio-demographic characteristics were ranked on the basis of their individual contributions to the total explained variance of income derived from the model. Also the differential effects of the categories of these characteristics were studied by means of the regression coefficients. Lastly, the age of head when the family's income would achieve its peak was determined by equating the first partial derivative of income with respect to the age of head to zero. Further, the income-age profiles were drawn by keeping the effects of all other variables fixed. All these empirical exercises were performed for the five regions and all Canada.

The major conclusions about socio-demographic characteristics, both in terms of their individual contributions to the total explained variance of income and their regression coefficients, at the regional level could be summed up as follows:

(i) We found that the head's main occupation in conjunction with his education was most effective in relation to the family's income over all the regions. Taking into account the significance of other characteristics, we observed that no two regions had identical rankings implying that these characteristics had different effects on the incomes of families living in different regions. For example, the proportion of earners ranked second in Quebec compared to its third position in Ontario and fourth in the remaining regions. On the other hand, the area of residence seemed more important in the Prairies compared to other characteristics; i.e., it ranked second in this region compared to its almost lowest position in the case of British Columbia. In Ontario, British Columbia and the Atlantic region, the education of head proved a significant factor (ranked second) in determining the family's income. Age of head and the wife's participation in labour force in conjunction with her education appeared to have greater influence on the family's income in British Columbia compared to all other regions.

(ii) A comparative study of estimated regression coefficients of the exogenous variables over the regions revealed certain interesting features:

(a) The inter-regional income differentials were low in the case of the skilled occupational groups. The variability between the estimated income differentials increased with a decrease in skill requirements of job or in the case of unskilled occupations. The inter-regional variation between incomes of families whose heads belonged to professional and technical group was the smallest.

(b) The regional income differentials decreased as level of family head's education increased.

(c) The inter-regional income dispersion with respect to the proportion of earners in a family was greater than the one with respect to the age of head.

(d) The inter-regional income differentials accruing to the wife's full-year participation in labour force increased with an increase in her level of education. This contrasted with the findings pertaining to the head's education as stated in (a) and (b).

(e) The inter-regional income dispersion in relation to area of residence was the smallest in the case of metropolitan areas and increased as we moved to other cities and small urban areas.

(f) Income differences with respect to the type of family (i.e. husband-wife family or other) were significant over all the regions.

At the national level the model identified the head's main occupation in conjunction with his education as the most important factor influencing the income of a Canadian family. Education of head and the proportion of earners in a family ranked in second and third place respectively. The influence of area of residence on the family's income was stronger than that of the regional location demonstrating the economic significance of the former. On the other hand, the contribution pertaining to the wife's participation in the labour force in conjunction with her education ranked in seventh position.

In terms of the individual variables we found that the three attributes influencing the total income of a Canadian family most were (i) the head having a university degree and being engaged in a professional or technical occupation, (ii) head holding a university degree and (iii) the proportion of earners in a family. Other significant attributes were the family's residence in a metropolitan area and living in Ontario; these were located in sixth and eighth position (out of 56) respectively.

There were two characteristics which did not perform according to our *a-priori* expectations. These included (a) the head's labour force participation and (b) his immigration status. In both cases, the effects of grouping may have weakened the estimates and their related statistical significance.

Limitations of the Model

Like most other econometric studies, this exposition has its weaknesses. The very first being the extensive reliance on dummy variables which in turn restricted the model to simple additivity. For example, specifying a quadratic or non-linear relationship between income and the head's participation in the labour force would have been more appropriate. The consequences of such over-simplification for operational reasons were

discernible in the case of categories pertaining to the head's participation in labour force throughout the analysis.

Secondly, there were problems arising from grouping of heterogeneous categories especially in the case of occupational and educational groups and labour force participation of head and wife. In some cases, these resulted in estimates that were difficult to interpret.

Thirdly, the analysis has omitted some pertinent characteristics which could have equally influenced the family's income. These would include information about the family's mobility, personal abilities of the head and other earners in a family, its background as well as other job related characteristics. Similarly, in the case of a wife, the information that was missing would include her age at marriage and above all her previous work experience.

Lastly, the model suffers from all the usual data deficiencies found in cross-sectional survey data—sampling and non-sampling errors.

Future Needs

One of the major needs is to increase the dimension of the analysis in terms of additional characteristics, some of which have been mentioned in the earlier paragraphs. This would enable us to explore various other issues which are equally pertinent in an income study like this. As the present study has been restricted to the analysis of total family income in relation to a specific number of socio-demographic characteristics of the family, this offered an over-all view about the incomes of families living in various regions of Canada.

Another extension of an analytical analysis would include a study of the individual components of income such as wages and salaries (before tax deductions), investment income and family and youth allowances. Such individual components of family income could then be explained by type and the size of family and its other socio-demographic characteristics.

It would also be interesting to study family income by type of its income recipients, as their contributions to the family's total income would depend on their socio-demographic characteristics. For example, the income contribution made by a young adult (16-25 years old) in a family and the one made by its head would more-or-less be a function of their individual characteristics such as education, work experience and number of weeks worked. Despite the fact that their explanatory characteristics are the same, their income contributions would be different; for example, the head reaping greater returns in the light of his longer work experience. Although our present study does not explain such income differences with respect to the types of income recipients, we believe the problem itself demands a more detailed analysis of incomes of families.

In addition, there is a need to study the cause-and-effects or feed-back relations between family income and its socio-demographic characteristics. This in turn would demand a change in the specification of the income model; i.e. from a single equation to a set of equations. In other words, the simultaneous equations model would be more useful in explaining the interdependent behaviour of income components and the characteristics of income recipients in the family.

APPENDICES

APPENDIX A

LIST OF VARIABLES

The following is a basic list of *variables* used in this study. The Roman letters without any suffix denote the broad sets of explanatory socio-demographic *characteristics* whereas their relative suffixes define the *categories* within such sets. For the purposes of estimation, these *categories* were transformed into dummy *variables*, i.e., by assigning the value 1 or zero, depending upon whether a given observation belonged to this *category* within the specified socio-demographic *characteristic* or not. In this manner, the regional and the national models contained 56 and 65 *variables* (including a constant term) respectively. As the applicability of principle of least squares required omission of one *variable* from each of the socio-demographic *characteristics*, the *variables* carrying asterics in the following list denote the reference *variables* used throughout the empirical analysis.

Y: Total actual income of a family in dollars

RG: Regional location of family

RG_1: Atlantic (*)

RG_2: Quebec

RG_3: Ontario

RG_4: Prairies

RG_5: British Columbia

AR: Area of residence of family

AR_1: Metropolitan (centers with a population of 30,000 and over)

AR_2: Other cities (centers with population 15,000 - 29,999)

AR_3: Small urban (centers with population under 15,000)

AR_4: Rural (*)

SH: Sex of head of family

SH_1: Male

SH_2: Female

MS: Marital status of the head of family

MS_1: Single

MS_2: Married

MS_3: Other (either a widow or a widower, legally separated or divorced)

EH: Education of the head of family

EH_1: No schooling (*)

EH_2: Some elementary school

EH_3: Finished elementary school

EH_4: Some high school

EH_5: Finished high school

EH_6: Some college or university

EH_7: University degree

ISH: Immigration status of the head of family

ISH_1: Born in Canada

ISH_2: Landed before 1946

ISH_3: Landed between 1946 to 1968 (*)

ISH_4: Not ascertained

EW: Education of wife

EW_1: No schooling

EW_2: Some elementary school

EW_3: Finished elementary school

EW_4: Some high school

EW_5: Finished high school

EW_6: Some college or university

EW_7: University degree

WLF: Wife's participation in labour force in 1967

WLF_1: Did not work

WLF_2: Worked 50 - 52 weeks a year, i.e. a full-year participation

WLF_3: Worked less than or equal to 49 weeks a year, i.e. a part-year participation.

HLF: Head's participation in labour force in 1967

HLFU: Did not work (*)

HLFE: Worked either full-year or part-year, i.e., participated in labour force (for the interpretation of the terms full-year and part-year, see definitions of WLF_2 and WLF_3 respectively)

MOH: Main occupation of the head of family during 1967 (determined by two considerations namely, (i) the greatest number of weeks spent on a specific job during a year and (ii) the kind of work performed during that period)

MOH_1: Did not work during 1967

MOH_2: Managerial

MOH_3: Professional and technical

MOH_4: Clerical

MOH_5: Sales

MOH_6: Service and recreation

MOH_7: Transport and communication

MOH_8: Farmers, farm workers, loggers and fishermen

MOH_9: Miners, quarrymen, craftsmen, production process and related workers

MOH_{10}: Labourers

NE/FS: Proportion of earners in a family derived by the ratio of total number of earners to the total size of family

AHF: Age of the head of family

AHF_1: Age in actual years

AHF_2: Square of AHF_1

In view of some technical and conceptual problems, some of these *variables* were transformed into composite *variables*, which in turn, were included in the final model. The following basic *characteristics* were transformed: SH, MS, WLF and MOH.

The necessity to create such composite *variables* arose after we found that the estimates pertaining to the socio-demographic *characteristics* MOH and WLF could not be interpreted in economic terms. On the other hand, in view of their significant roles in determining the family's income, these could not be discarded from the model. Therefore, we decided to study these *characteristics* in conjunction with certain other *characteristics*; that is, MOH in conjunction with EH and WLF in conjunction with EW.[1] The corresponding *categories* of these cross-classified *characteristics* as specified in the model, have been defined as follows:

MOH.EH: Main occupation of the head of family in conjunction with his education.

MOH_{21}: Managerial with some or finished elementary education

MOH_{22}: Managerial with some or finished high school education

MOH_{23}: Managerial with some college or university education

MOH_{24}: Managerial with a university degree

MOH_{31}: Professional and technical with some or finished elementary education

MOH_{32}: Professional and technical with some or finished high school education

MOH_{33}: Professional and technical with some college or university education

MOH_{34}: Professional and technical with a university degree

MOH_{41}: Clerical with some or finished elementary education

MOH_{42}: Clerical with some or finished high school education

MOH_{43}: Clerical with some or finished university education

MOH_{51}: Sales with some or finished elementary education

MOH_{52}: Sales with some or finished high school education

MOH_{53}: Sales with some or finished university education

MOH_{61}: Service and recreation with some or finished elementary education

MOH_{62}: Service and recreation with some or finished high school education

MOH_{63}: Service and recreation with some or finished university education

MOH_{71}: Transport and communication with some or finished elementary education

MOH_{72}: Transport and communication with some or finished high school education

MOH_{73}: Transport and communication with some or finished university education

MOH_{81}: Farmers, farm workers, loggers, etc. with some or finished elementary education

MOH_{82}: Farmers, farm workers, loggers, etc. with more than elementary education

MOH_{91}: Miners, craftsmen, production process and related workers with some or finished elementary education

MOH_{92}: Miners, craftsmen, production process and related workers with some or finished high school education

MOH_{93}: Miners, craftsmen, production process and related workers with some or finished university education

MOH_{101}: Labourers with some or finished elementary education

MOH_{102}: Labourers with more than elementary education

MOH(R): The residual group consisting of those who did not work during 1967 and also had no schooling (*)

WLF.EW: Wife's participation in labour force in conjunction with her education

WLF_{21}: Full-year participation with some or finished elementary education

WLF_{22}: Full-year participation with some or finished high school education

WLF_{23}: Full-year participation with some college or university education

WLF_{24}: Full-year participation with a university degree

[1] See Appendix B for details about the selection of cross-classified *characteristics*.

WLF_{31}: Part-year participation with some or finished elementary education

WLF_{32}: Part-year participation with some or finished high school education

WLF_{33}: Part-year participation with some college or university education

WLF_{34}: Part-year participation with a university degree

The composite *variables* MOH.EH and WLF.EW were created by multiplying the column vectors of individual *categories* or the column vectors obtained by merging two or three *categories* of one or the other exogenous *characteristics*. For example, MOH_{21}, MOH_{22}, MOH_{43}, etc., were derived from the expressions $MOH_2.(EH_2 + EH_3)$, $MOH_2.(EH_4 + EH_5)$, ... $MOH_4.(EH_6 + EH_7)$ respectively. Similarly, WLF_{21}, WLF_{22}, ... etc., were obtained by the product terms $WLF_2.(EW_2 + EW_3)$, $WLF_2.(EW_4 + EW_5)$, ..., respectively.

As regards the specification of WLF in conjunction with EW was concerned, the nature of the model called for a further refinement in its definition. Because as it stood, WLF.EW defined the wife's participation in labour force in conjunction with her education irrespective of the type of economic family and there were cases of economic families with no wives present. These included in broad terms, families headed by (i) females, (ii) single males and (iii) males with marital status MS_3. So in view of the fact that WLF.EW would be relevant only in the case of husband-wife families, we had to use WLF.EW in conjunction with a new composite *characteristic*-NF_1, identifying families with married males as heads. Thus for the purposes of estimation, we included ($WLF.EW.NF_1$), denoted by WLF+ in the main model, and the resulting coefficients pertaining to its *categories* (or the corresponding dummy variables) depicted the income differences accruing to different employment patterns of wives in husband-wife families. The rest of the economic families with female heads and those with male heads with marital status MS_3 were identified by a composite *characteristic*-NF_2 created by the arithmetic operation:

$$[\{(SH_1) \times (MS_1)\} + \{(SH_1) \times (MS_3)\} + (SH_2)]$$

where (SH_1), (MS_1), (MS_3) and (SH_2) denoted the column vectors with their definitions as described in the basic list. The estimated coefficient of NF_2 in the model depicted the income difference between a "husband-wife" family and the "others" (i.e. families with no wives) as NF_1 was treated as a reference *variable*.

Finally, the *categories* of WLF+ as used in the model have the following interpretations:

WLF+: Wife's participation in labour force in conjunction with her education in the case of a husband-wife family

$WLF+_{21}$: Full-year participation with some or finished elementary education

$WLF+_{22}$: Full-year participation with some or finished high school education

$WLF+_{23}$: Full-year participation with some college or university education

$WLF+_{24}$: Full-year participation with a university degree

$WLF+_{31}$: Part-year participation with some or finished elementary education

$WLF+_{32}$: Part-year participation with some or finished high school education

$WLF+_{33}$: Part-year participation with some college or university education

$WLF+_{34}$: Part-year participation with a university degree

WLF+(R): The residual group consisting of wives who did not work and also had no schooling (*)

Other statistical notations used in this study have been as follows:

$X_{i,j}$: The j-th *variable* within the i-th explanatory *characteristic* (broad qualitative or quantitative). The suffix j takes the values 1 to 5 for RG, 1 to 4 for AR, and ISH, 1 to 7 for EH, 1 to 28 for MOH.EH, 1 to 9 for WLF+ and 1 to 2 for HLF and NF respectively. Similarly, the suffix i takes the values 1 to 8 in a regional model compared to 1 to 10 in the national model.

$\bar{X}_{i,j}$: Mean of the j-th *variable* within the i-th explanatory *characteristic* (broad qualitative or quantitative).

$\sigma_{\bar{X}_{i,j}}$: Standard deviation of the j-th *variable* within the i-th explanatory *characteristic* (broad qualitative or quantitative).

$b_{i,j}$: Unknown parameter of the j-th *variable* within the i-th explanatory *characteristic* (broad qualitative or quantitative).

$\hat{b}_{i,j}$: Least squares estimate of $b_{i,j}$, depicting the respective income differential with respect to a given reference *variable*.

$S(\hat{b}_{i,j})$: Standard error of $\hat{b}_{i,j}$.

t-statistic: Defined as a ratio of $\hat{b}_{i,j}/S(\hat{b}_{i,j})$. In view of the fact that the regional sample sizes were large, this t-statistic in theoretical terms approached a normal distribution. Hence the value of this ratio exceeding 1.96 depicted statistical significance of the given *variable* in question. However, as we have been mostly using dummy *variables*, their statistical significance so derived, in fact, reflected the significance of the differences between the given *variables* and their corresponding reference *variables*.

α: Level of significance or the Type-I error, which has been assigned the 5% value throughout the analysis in order to keep our inferences quite consistent with the well established and conventional practices.

C.V.: Coefficient of variation, defined as a ratio of standard deviation to the mean, has been used to measure the variability between the sets of estimates.

F: Conventional F-statistic testing the significance of variances and the goodness-of-fit etc.

APPENDIX B

SOME MAJOR PROBLEMS OF SPECIFICATION

This appendix outlines certain major problems of specification we encountered in the process of postulating the income model described here. While defining and resolving these technical problems we have refrained from augmenting our arguments with any empirical results. In other words, their exposition is purely descriptive in nature.

We began this econometric exercise by postulating total income of a family as a function of its socio-demographic characteristics. In the case of a regional model, these included (i) area of residence; (ii) sex of head; (iii) education of head; (iv) number of children (a) below 6 years of age, (b) 6-15 years and (c) between 16-25 years; (v) age of head; (vi) number of weeks the head worked during the year; (vii) wife's participation in labour force; (viii) main occupation of head; and (ix) the proportion of earners in a family. The characteristics (v) and (vi), moreover, were represented in terms of groups; i.e., the heads were grouped into six age groups, namely (a) below 24 years, (b) 25-34 years, (c) 35-44 years, (d) 45-54 years, (e) 55-64 years and (f) 65 years and over and the weeks worked into seven groups such as (1) 50-52 weeks, (2) 40-49 weeks, (3) 30-39 weeks, (4) 20-29 weeks, (5) 10-19 weeks, (6) 1-9 weeks and (7) did not work. One of the considerations which motivated us to specify number of weeks worked in such groupings was to study the effects of break-downs in the working pattern of head on the family's income.

The resulting estimates pertaining to the age of head, wife's participation in labour force, and main occupation of head were not only lacking meaningful interpretation but also depicted certain inconsistencies. For example, the age variable showed that the family's income was highest in the age group 65 years and over whereas the truth was that the income achieved its peak by the time the head had reached the early or mid-fifties and then started to decline gradually. Therefore, such conflicting inferences pertaining to the above-mentioned characteristics forced us to seek some suitable alternatives about their specification.

In view of their individual pertinence to the family's income, it was quite inconceivable to exclude them from the model simply because they were statistically inconsistent. On the other hand, the alternatives were (i) the different specification and (ii) their specification in conjunction with some other characteristics.

We introduced the age variable in terms of actual number of years and also its squared term in the model in order to demonstrate a quadratic relationship between income and age. The objective was to evaluate the optimal point for the income-age curve (keeping the effects of other variables constant) by equating the first partial derivative of income with respect to age to zero.

As regards the main occupation of head, the choice was to specify this in conjunction with either the number of weeks worked, the head's age or his education. Of course, such a choice was conditioned by certain realities. For example, the magnitude of financial returns to certain occupations such as construction, logging, fishing and farming is highly influenced by the seasonal factors affecting their employment or in other words, the number of weeks worked. But in view of the broad groupings of occupations, we eventually decided to choose that particular characteristic which would equally explain the income progressions within a particular occupational category. In this respect, one's skill, experience, and other abilities vital to earning capability must be taken into account. As education alone influences these personal characteristics, we decided to specify the main occupation of head in conjunction with his education in the final version of the model.

Similarly, the wife's participation in the labour force was studied in relation to her age, number of children by age groups, size of family and her education. The resulting empirical research demonstrated the statistical superiority (in terms of plausible results) of the cross-classified group namely, the wife's participation in labour force in conjunction with her education. Moreover, education of the wife may in fact, determine her taste for market work.

The selection of the corresponding categories for the above listed cross-classified characteristics, was conditioned by a study of the cells of their tables. The cells with low or no frequencies corresponding to either of the characteristics were merged in order to reduce the total number of categories. For example, with the exception of managerial, professional and technical, farmers, farm workers and fishermen, all other occupational categories were taken in conjunction with three broad categories of education, namely, some or finished elementary schooling, some or finished high schooling and some university education or a university degree. The reference group in this case consisted of all those who did not work and had no education. A similar description holds for the other cross-classified characteristics, the reference group of which contained all wives who did not participate in the labour force and had no education.

We re-estimated the model after incorporating these changes in the initial specification. It was rather appalling to find the optimal age of head exceeding 100 years and estimates pertaining to the number of weeks worked that could not be interpreted. This added a new dimension to the problem of specification.

At this juncture we studied the matrix of simple correlations (for the region in question) and especially the correlations between age of

head, number of weeks worked and the number of children by age groups. These correlations showed a good degree of relationship between the head's age and the number of children and the head's age and the number of weeks worked. As regards the former, there was no problem about accepting the validity of such correlations because in real terms, the ages of children do reflect the age of the head of family. But the revelation about the relationship between the age and number of weeks worked created a serious problem. Besides testing these correlations, we also employed certain other statistical tests to recheck the authenticity of this situation. Coincidently, these tests confirmed the existence of relationship between age and other characteristics.

It became obvious at this stage that the interrelationship between these socio-demographic characteristics could not be explored by using a single equation model. For example, there exists a parabolic relationship between age and number of weeks worked; i.e., number of weeks worked would increase with an increase in age up to certain point beyond which they would begin to decline as one got older.

Thus in the light of existing constraints about the model, we rectified this problem of specification by excluding both the number of children (by age groups) and the number of weeks worked (by groups) from the model. On the other hand, in view of the fact that the working pattern of head, i.e., full-year or part-year participation in the labour force, affected the family's income, we had had to find a substitute for the number of weeks worked, as described above. To this end, we specified two broad groups, namely, either the head worked or he did not work, in the final version of the model. Obviously such broad groupings of working patterns would create certain statistical problems. This has been the basic deficiency of our model.

Lastly, the concept of the economic family created certain specification problems. An economic family, by definition, may or may not have a wife as its member. Therefore, the specification of cross-classified characteristics defining the wife's participation in labour force in conjunction with her education needed some modification so that the final estimates pertaining to this group would be relevant only in the case of husband-wife families. To achieve this purpose, we defined a new characteristic, namely the type of family. This classified husband-wife families (headed by males) and single parent and all other types of families (headed by either a male or female). Such a differentiation of families, in turn, enabled us to exclude the sex of head as an exogenous variable from the final version of the model. Finally, we defined certain composite variables pertaining to the wife's participation in labour force in conjunction with her education that were valid only for the husband-wife families. The details of this procedure are given in Appendix A.

Thus, the final version of the regional model presented in this text explains income as a function of (i) area of residence of family; (ii) education of head; (iii) main occupation of head in conjunction with his education; (iv) wife's participation in labour force in conjunction with her education; (v) head's participation in labour force; (vi) his age; (vii) the proportion of earners in a family and (viii) the type of family. A comparison of the initial and the final versions of the model thus describes the evolution in the process of specification and also gives an idea about the magnitude of the empirical work involved in this exercise.

APPENDIX C

REGRESSION RESULTS

TABLE C.1. Parameter Estimates of Model for the Atlantic Region[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
AR_1	0.2812	0.4496	1,447.50	11.25
AR_2	0.1472	0.3543	1,085.48	6.95
AR_3	0.1184	0.3231	315.26	1.88
AR_4	0.4532	...	—	—
EH_1	0.0315	...	—	—
EH_2	0.3281	0.4695	- 235.59	- 0.64
EH_3	0.1747	0.3797	290.54	0.76
EH_4	0.2677	0.4428	254.42	0.60
EH_5	0.1150	0.3190	1,330.77	3.00
EH_6	0.0467	0.2111	1,108.53	1.63
EH_7	0.0363	0.1871	2,627.78	3.06
MOH(R)	0.1884	...	—	—
MOH_{21}	0.0184	0.1344	1,434.19	2.71
MOH_{22}	0.0564	0.2305	2,589.53	5.31
MOH_{23}	0.0099	0.0985	4,303.84	5.03
MOH_{24}	0.0044	0.0658	5,060.21	4.39
MOH_{31}	0.0016	0.0388	2,476.37	1.84
MOH_{32}	0.0234	0.1511	2,131.35	3.85
MOH_{33}	0.0141	0.1173	2,404.82	2.99
MOH_{34}	0.0272	0.1623	4,209.84	4.59
MOH_{41}	0.0072	0.0839	1,647.46	2.30
MOH_{42}	0.0335	0.1800	1,134.30	2.20
MOH_{43}	0.0071	0.0834	1,032.84	1.15
MOH_{51}	0.0077	0.0868	1,463.29	2.10
MOH_{52}	0.0213	0.1442	1,529.87	2.76
MOH_{53}	0.0026	0.0500	3,213.01	2.67
MOH_{61}	0.0356	0.1851	335.80	0.71
MOH_{62}	0.0242	0.1532	460.43	0.85
MOH_{63}	0.0026	0.0497	555.39	0.46
MOH_{71}	0.0421	0.2006	1,191.19	2.61
MOH_{72}	0.0277	0.1642	1,584.26	2.97
MOH_{73}	0.0013	0.0364	3,042.74	2.00
MOH_{81}	0.0727	0.2595	218.35	0.51
MOH_{82}	0.0295	0.1692	66.21	0.13
MOH_{91}	0.1679	0.3738	1,183.42	2.91
MOH_{92}	0.1193	0.3241	1,582.81	3.41
MOH_{93}	0.0058	0.0751	2,602.68	2.76
MOH_{101}	0.0366	0.1874	501.55	1.07
MOH_{102}	0.0115	0.1066	- 13.71	- 0.02
$WLF^+(R)$	0.7498	...	—	—
WLF^+_{21}	0.0425	0.2017	- 440.38	- 1.64
WLF^+_{22}	0.1089	0.3115	501.59	2.56
WLF^+_{23}	0.0181	0.1333	1,248.71	3.13
WLF^+_{24}	0.0057	0.0755	738.26	1.08
WLF^+_{31}	0.0233	0.1508	- 1,133.63	- 3.33
WLF^+_{32}	0.0466	0.2109	- 59.45	- 0.23
WLF^+_{33}	0.0030	0.0546	- 1,000.57	- 1.08
WLF^+_{34}	0.0021	0.0455	1,940.12	1.72
HLFE	0.8112	0.3913	42.82	0.12
HLFU	0.1888	...	—	—
AHF_1	46.95	15.46	188.02	9.37
AHF_2	2,443.46	1,563.69	- 1.60	- 7.81
NE/FS	0.3891	0.2700	2,243.03	9.04
NF_1	0.8850	...	—	—
NF_2	0.1150	0.3190	- 857.93	- 4.70
Constant	...	...	- 2,046.04	- 3.59

[1] See Appendix A for the definitions of statistical notations used in the table.

TABLE C.2. Parameter Estimates of Model for Quebec[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
AR_1	0.6465	0.4781	1,155.32	5.32
AR_2	0.0468	0.2112	584.85	1.57
AR_3	0.1237	0.3292	314.90	1.18
AR_4	0.1830	...	–	–
EH_1	0.0138	...	–	–
EH_2	0.2901	0.4538	528.37	0.76
EH_3	0.1902	0.3925	837.55	1.18
EH_4	0.2292	0.4203	598.18	0.77
EH_5	0.1628	0.3692	1,420.52	1.81
EH_6	0.0432	0.2033	700.88	0.62
EH_7	0.0707	0.2562	2,697.15	2.35
MOH(R)	0.1479	...	–	–
MOH_{21}	0.0236	0.1520	1,313.33	1.79
MOH_{22}	0.0550	0.2279	3,961.83	5.38
MOH_{23}	0.0095	0.0972	5,747.59	4.48
MOH_{24}	0.0138	0.1168	6,261.55	5.10
MOH_{31}	0.0029	0.0538	424.13	0.30
MOH_{32}	0.0344	0.1822	2,646.54	3.38
MOH_{33}	0.0120	0.1088	3,754.01	3.03
MOH_{34}	0.0462	0.2100	5,013.27	4.46
MOH_{41}	0.0146	0.1199	632.92	0.76
MOH_{42}	0.0398	0.1954	1,014.36	1.34
MOH_{43}	0.0062	0.0784	1,518.90	1.13
MOH_{51}	0.0075	0.0864	1,850.78	1.86
MOH_{52}	0.0252	0.1567	1,952.43	2.42
MOH_{53}	0.0079	0.0884	2,841.57	2.25
MOH_{61}	0.0402	0.1964	– 405.86	– 0.60
MOH_{62}	0.0318	0.1754	1,073.88	1.39
MOH_{63}	0.0024	0.0489	– 401.29	– 0.23
MOH_{71}	0.0364	0.1873	677.55	0.97
MOH_{72}	0.0265	0.1606	1,007.17	1.25
MOH_{73}	0.0015	0.0393	4,386.01	2.18
MOH_{81}	0.0531	0.2243	– 154.72	– 0.23
MOH_{82}	0.0084	0.0910	– 218.67	– 0.22
MOH_{91}	0.1698	0.3755	765.14	1.24
MOH_{92}	0.1325	0.3391	1,384.60	1.97
MOH_{93}	0.0074	0.0859	2,260.18	1.78
MOH_{101}	0.0339	0.1809	432.27	0.62
MOH_{102}	0.0096	0.0973	48.19	0.05
$WLF^+(R)$	0.8386	...	–	–
WLF^+_{21}	0.0636	0.2440	– 281.73	– 0.91
WLF^+_{22}	0.0100	0.2999	431.24	1.54
WLF^+_{23}	0.0091	0.0949	277.64	0.37
WLF^+_{24}	0.0084	0.0912	361.53	0.46
WLF^+_{31}	0.0278	0.1644	– 1,059.21	– 2.46
WLF^+_{32}	0.0361	0.1865	– 1,041.02	– 2.74
WLF^+_{33}	0.0034	0.0585	– 3,742.19	– 3.16
WLF^+_{34}	0.0030	0.0545	– 823.17	– 0.65
HLFE	0.8452	0.3617	710.03	1.36
HLFU	0.1548	...	–	–
AHF_1	44.65	14.40	326.98	10.90
AHF_2	2,201.20	1,421.04	– 2.87	– 9.20
NE/FS	0.4103	0.2758	3,350.58	10.14
NF_1	0.8957	...	–	–
NF_2	0.1043	0.3056	– 781.71	– 2.92
Constant	...	...	– 5,748.21	– 6.23

[1] See footnote to Table C.1.

TABLE C.3. Parameter Estimates of Model for Ontario[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
AR_1	0.7027	0.4571	1,387.51	6.05
AR_2	0.0585	0.2347	269.16	0.75
AR_3	0.0947	0.2928	515.99	1.68
AR_4	0.1441	...	–	–
EH_1	0.0052	...	–	–
EH_2	0.1216	0.3268	– 160.43	– 0.15
EH_3	0.2428	0.4288	368.81	0.35
EH_4	0.3247	0.4683	596.60	0.55
EH_5	0.1783	0.3827	900.28	0.83
EH_6	0.0482	0.2141	1,033.04	0.77
EH_7	0.0792	0.2701	3,722.27	2.81
MOH(R)	0.1210	...	–	–
MOH_{21}	0.0175	0.1310	5,281.78	5.98
MOH_{22}	0.0774	0.2673	5,172.06	6.58
MOH_{23}	0.0118	0.1082	5,588.04	4.47
MOH_{24}	0.0145	0.1195	9,799.35	8.03
MOH_{31}	0.0030	0.0548	1,931.25	1.31
MOH_{32}	0.0390	0.1937	3,424.90	4.21
MOH_{33}	0.0171	0.1296	4,056.28	3.36
MOH_{34}	0.0527	0.2235	5,790.32	5.22
MOH_{41}	0.0125	0.1113	1,580.80	1.66
MOH_{42}	0.0490	0.2158	1,970.49	2.45
MOH_{43}	0.0054	0.0731	2,045.44	1.49
MOH_{51}	0.0072	0.0843	2,093.18	1.91
MOH_{52}	0.0322	0.1766	3,720.42	4.46
MOH_{53}	0.0048	0.0694	3,218.89	2.28
MOH_{61}	0.0296	0.1695	799.30	0.99
MOH_{62}	0.0317	0.1753	1,885.19	2.29
MOH_{63}	0.0023	0.0477	4,368.92	2.42
MOH_{71}	0.0231	0.1502	2,075.94	2.44
MOH_{72}	0.0301	0.1707	2,476.05	2.90
MOH_{73}	0.0007	0.0261	3,671.90	1.26
MOH_{81}	0.0288	0.1673	406.71	0.50
MOH_{82}	0.0216	0.1454	153.04	0.17
MOH_{91}	0.1488	0.3559	2,052.70	2.82
MOH_{92}	0.1780	0.3825	2,761.76	3.63
MOH_{93}	0.0084	0.0915	2,675.57	2.11
MOH_{101}	0.0210	0.1433	1,428.12	1.69
MOH_{102}	0.0108	0.1032	1,456.27	1.44
WLF+(R)	0.6331	...	–	–
$WLF+_{21}$	0.0516	0.2212	232.48	0.66
$WLF+_{22}$	0.1773	0.3819	372.75	1.50
$WLF+_{23}$	0.0180	0.1330	893.47	1.58
$WLF+_{24}$	0.0106	0.1026	605.36	0.82
$WLF+_{31}$	0.0257	0.1583	– 1,631.12	– 3.50
$WLF+_{32}$	0.0733	0.2606	– 792.15	– 2.69
$WLF+_{33}$	0.0057	0.0750	– 1,048.94	– 1.10
$WLF+_{34}$	0.0047	0.0685	440.85	0.42
HLFE	0.8685	0.3379	– 95.86	– 0.16
HLFU	0.1315	...	–	–
AHF_1	45.60	15.10	321.56	10.71
AHF_2	2,307.20	1,505.25	– 2.85	– 9.04
NE/FS	0.4693	0.2932	2,592.27	7.52
NF_1	0.9104	...	–	–
NF_2	0.0896	0.2856	– 1,343.77	– 4.65
Constant	...	...	– 5,062.79	– 4.16

[1] See footnote to Table C.1.

TABLE C.4. Parameter Estimates of Model for the Prairie Region[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{X_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
AR_1	0.5064	0.5000	1,801.57	7.91
AR_2	0.0430	0.2028	1,095.29	2.84
AR_3	0.1698	0.3755	624.55	2.44
AR_4	0.2808	...	–	–
EH_1	0.0148	...	–	–
EH_2	0.1743	0.3794	- 214.39	- 0.33
EH_3	0.2386	0.4262	242.42	0.37
EH_4	0.2965	0.4567	436.20	0.61
EH_5	0.1721	0.3774	789.83	1.08
EH_6	0.0475	0.2127	76.55	0.08
EH_7	0.0563	0.2305	1,529.38	1.35
MOH(R)	0.1338	...	–	–
MOH_{21}	0.0209	0.1431	3,704.01	4.78
MOH_{22}	0.0704	0.2557	4,063.66	5.45
MOH_{23}	0.0101	0.1000	5,546.35	4.66
MOH_{24}	0.0077	0.0876	9,726.10	7.19
MOH_{31}	0.0025	0.0502	1,647.53	1.09
MOH_{32}	0.0261	0.1594	2,599.99	3.15
MOH_{33}	0.0144	0.1190	3,788.98	3.40
MOH_{34}	0.0413	0.1990	5,438.94	4.67
MOH_{41}	0.0077	0.0874	2,226.98	2.23
MOH_{42}	0.0358	0.1857	1,509.45	1.94
MOH_{43}	0.0053	0.0727	3,364.62	2.55
MOH_{51}	0.0086	0.0922	684.39	0.71
MOH_{52}	0.0343	0.1820	2,605.39	3.31
MOH_{53}	0.0052	0.0722	3,638.07	2.77
MOH_{61}	0.0236	0.1519	295.79	0.39
MOH_{62}	0.0320	0.1760	1,385.75	1.75
MOH_{63}	0.0025	0.0498	2,219.80	1.34
MOH_{71}	0.0242	0.1536	1,502.01	1.97
MOH_{72}	0.0402	0.1965	2,080.81	2.66
MOH_{73}	0.0014	0.0380	1,117.48	0.55
MOH_{81}	0.1206	0.3256	589.37	0.90
MOH_{82}	0.0770	0.2666	1,252.50	1.69
MOH_{91}	0.1021	0.3028	2,015.37	3.08
MOH_{92}	0.1224	0.3277	2,376.40	3.29
MOH_{93}	0.0054	0.0731	3,057.31	2.31
MOH_{101}	0.0158	0.1247	204.23	0.24
MOH_{102}	0.0087	0.0927	780.43	0.77
WLF+(R)	0.6398	...	–	–
$WLF+_{21}$	0.0475	0.2127	- 195.95	- 0.55
$WLF+_{22}$	0.1418	0.3489	364.59	1.46
$WLF+_{23}$	0.0207	0.1424	2,330.95	4.55
$WLF+_{24}$	0.0091	0.0950	1,375.20	1.81
$WLF+_{31}$	0.0373	0.1895	- 882.14	- 2.33
$WLF+_{32}$	0.0909	0.2875	- 564.67	- 2.15
$WLF+_{33}$	0.0094	0.0965	- 1,081.27	- 1.49
$WLF+_{34}$	0.0035	0.0595	903.64	0.76
HLFE	0.8570	0.3501	- 97.83	- 0.18
HLFU	0.1430	...	–	–
AHF_1	46.88	15.51	283.23	9.76
AHF_2	2,438.69	1,570.24	- 2.64	- 8.80
NE/FS	0.4560	0.2887	1,997.38	6.22
NF_1	0.9079	...	–	–
NF_2	0.0921	0.2891	- 1,455.13	- 5.23
Constant	...	...	- 4,002.93	- 4.58

[1] See footnote to Table C.1.

TABLE C.5. Parameter Estimates of Model for British Columbia[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
AR_1	0 5930	0 4913	1,053.16	3 75
AR_2	0.1335	0.3401	1,190.70	3.37
AR_3	0.1261	0.3320	567.88	1.58
AR_4	0.1474	...	–	–
EH_1	0.0063	...	–	–
EH_2	0.1040	0.3053	386.28	0.30
EH_3	0.2015	0.4011	431.13	0.33
EH_4	0.3329	0.4713	658.22	0.51
EH_5	0.2073	0.4054	652.12	0.50
EH_6	0.0825	0.2751	2,072.28	1.36
EH_7	0.0655	0.2475	3,260.07	1.96
MOH(R)	0.1526	...	–	–
MOH_{21}	0.0237	0.1522	3,156.28	2.58
MOH_{22}	0.0889	0.2846	4,648.20	4.21
MOH_{23}	0.0165	0.1275	4,987.01	3.26
MOH_{24}	0.0076	0.0866	5,881.60	3.26
MOH_{31}	0.0016	0.0396	5,972.53	2.34
MOH_{32}	0.0202	0.1407	3,322.64	2.65
MOH_{33}	0.0237	0.1520	2,138.20	1.44
MOH_{34}	0.0477	0.2131	4,936.57	3.13
MOH_{41}	0.0090	0.0944	1,429.09	0.98
MOH_{42}	0.0402	0.1965	2,852.59	2.43
MOH_{43}	0.0077	0.0874	1,322.49	0.78
MOH_{51}	0.0036	0.0602	885.03	0.48
MOH_{52}	0.0355	0.1850	3,353.34	2.89
MOH_{53}	0.0088	0.0934	2,376.60	1.44
MOH_{61}	0.0237	0.1522	797.61	0.66
MOH_{62}	0.0380	0.1912	1,265.57	1.08
MOH_{63}	0.0051	0.0709	2,387.53	1.30
MOH_{71}	0.0195	0.1384	2,371.42	1.91
MOH_{72}	0.0414	0.1992	3,054.72	2.60
MOH_{73}	0.0030	0.0549	608.14	0.28
MOH_{81}	0.0313	0.1741	1,087.07	0.93
MOH_{82}	0.0255	0.1577	1,282.03	1.06
MOH_{91}	0.0992	0.2989	2,530.56	2.28
MOH_{92}	0.1807	0.3848	3,192.63	2.90
MOH_{93}	0.0159	0.1250	2,822.56	1.85
MOH_{101}	0.0140	0.1173	1,107.42	0.83
MOH_{102}	0.0154	0.1232	2,089.02	1.60
WLF+(R)	0.6462	...	–	–
$WLF+_{21}$	0.0286	0.1668	– 99.72	– 0.17
$WLF+_{22}$	0.1571	0.3639	334.02	1.00
$WLF+_{23}$	0.0280	0.1648	2,778.84	4.54
$WLF+_{24}$	0.0102	0.1005	2,225.50	2.35
$WLF+_{31}$	0.0244	0.1543	– 822.28	– 1.32
$WLF+_{32}$	0.0932	0.2908	– 99.78	– 0.29
$WLF+_{33}$	0.0112	0.1054	1,935.35	2.19
$WLF+_{34}$	0.0011	0.3265	– 5,526.59	– 1.96
HLFE	0.8426	0.3642	– 222.84	– 0.23
HLFU	0.1574	...	–	–
AHF_1	46.48	15.80	293.73	8.15
AHF_2	2,410.51	1,616.29	– 2.83	– 7.60
NE/FS	0.4559	0.2973	1,932.97	4.18
NF_1	0.9151	...	–	–
NF_2	0.0849	0.2787	– 1,436.41	– 3.91
Constant	...	...	– 3,938.91	– 2.69

[1] See footnote to Table C.1.

TABLE C.6. Parameter Estimates of Model for All Canada[1]

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
RG_1	0.0865	...	–	–
RG_2	0.2757	0.4469	796.30	6.09
RG_3	0.3677	0.4822	1,311.64	10.13
RG_4	0.1685	0.2290	473.59	2.59
RG_5	0.1016	0.3021	870.15	5.59
AR_1	0.6065	0.4885	1,411.59	14.00
AR_2	0.0679	0.2517	794.63	5.23
AR_3	0.1206	0.3256	440.79	3.49
AR_4	0.2050	...	–	–
EH_1	0.0117	...	–	–
EH_2	0.1930	0.3947	- 16.04	- 0.05
EH_3	0.2175	0.4126	284.63	0.80
EH_4	0.2895	0.4536	282.09	0.76
EH_5	0.1704	0.3760	762.68	2.02
EH_6	0.0500	0.2180	781.90	1.54
EH_7	0.0679	0.2516	2,747.30	5.17
ISH_1	0.5548	0.4970	555.25	4.58
ISH_2	0.0752	0.2637	338.07	1.98
ISH_3	0.0919	...	–	–
ISH_4	0.2781	0.4481	357.79	2.79
MOH(R)	0.1398	...	–	–
MOH_{21}	0.0205	0.1416	3,119.90	8.30
MOH_{22}	0.0694	0.2541	4,354.38	12.45
MOH_{23}	0.0112	0.1053	5,292.21	9.32
MOH_{24}	0.0116	0.1070	8,155.28	13.97
MOH_{31}	0.0026	0.0512	1,574.23	2.23
MOH_{32}	0.0323	0.1768	2,948.83	7.86
MOH_{33}	0.0156	0.1240	3,380.38	6.22
MOH_{34}	0.0463	0.2101	5,289.32	10.11
MOH_{41}	0.0115	0.1064	1,207.61	2.79
MOH_{42}	0.0420	0.2006	1,595.50	4.40
MOH_{43}	0.0060	0.0770	1,839.53	2.99
MOH_{51}	0.0072	0.0844	1,486.32	3.03
MOH_{52}	0.0300	0.1706	2,774.81	7.38
MOH_{53}	0.0059	0.0769	2,922.34	4.76
MOH_{61}	0.0314	0.1745	263.69	0.75
MOH_{62}	0.0318	0.1754	1,330.60	3.57
MOH_{63}	0.0026	0.0513	2,021.74	2.60
MOH_{71}	0.0282	0.1656	1,379.70	3.83
MOH_{72}	0.0317	0.1753	1,993.54	5.27
MOH_{73}	0.0013	0.0366	2,783.89	2.78
MOH_{81}	0.0550	0.2280	241.13	0.72
MOH_{82}	0.0284	0.1660	557.48	1.46
MOH_{91}	0.1433	0.3504	1,532.47	4.88
MOH_{92}	0.1513	0.3583	2,274.59	6.73
MOH_{93}	0.0082	0.0900	2,662.63	4.61
MOH_{101}	0.0243	0.1539	774.73	2.12
MOH_{102}	0.0106	0.1025	877.62	1.93
$WLF^+(R)$	0.6776	...	–	–
WLF^+_{21}	0.0511	0.2208	- 22.08	- 0.14
WLF^+_{22}	0.1420	0.3491	371.65	3.10
WLF^+_{23}	0.0170	0.1294	1,409.48	5.28
WLF^+_{24}	0.0093	0.0959	820.44	2.30
WLF^+_{31}	0.0279	0.1647	- 1,203.27	- 5.86
WLF^+_{32}	0.0657	0.2478	- 632.82	- 4.48
WLF^+_{33}	0.0060	0.0773	- 1,067.94	- 2.51
WLF^+_{34}	0.0034	0.0585	91.45	0.16

[1] See footnote to Table C.1.

TABLE C.6. Parameter Estimates of Model for All Canada[1] – Concluded

$X_{i,j}$	$\bar{X}_{i,j}$	$\sigma_{\bar{X}_{i,j}}$	$\hat{b}_{i,j}$	t-statistics
HLFE	0.8526	0.3546	134.90	0.51
HLFU	0.1474	...	-	
AHF_1	45.76	15.11	306.92	22.51
AHF_2	2,322.42	1,513.52	- 2.78	- 19.55
NE/FS	0.4425	0.2877	2,624.35	16.74
NF_1	0.9042	...	–	–
NF_2	0.0958	0.2943	- 1,132.89	- 8.79
Constant	...	...	- 5,851.44	- 12.42

[1] See footnote to Table C.1.

BIBLIOGRAPHY

(1) ADAMS, F.G. (1958), The size of individual incomes: socio-economic variables and chance variation, *Review of Economics and Statistics*, 40, 390-98.

(2) AITCHISON, J. and BROWN, J.A.C. (1957), *The Log-normal Distribution* (with special reference to its uses in economics), Cambridge: Cambridge University Press.

(3) BECKER, G.S. (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York: Columbia University Press.

(4) BECKER, G.S. and CHISWICK, B.R. (1966), Education and the distribution of earnings, *American Economic Review* (Papers and Proceedings), 56, 358-69.

(5) BELLOC, N.B. (1950), Labour force participation and employment opportunities of women, *Journal of American Statistical Association*, 45, 401-10.

(6) BJERKE, K. (1970), Income and wage distributions, Part I: A survey of the literature, *Review of Income and Wealth*, No. 3, 235-51.

(7) BLAU, P.M. and DUNCAN, O.D. (1967), *The American Occupational Structure*, New York: John Wiley & Sons.

(8) BOWEN, W.G. and FINEGAN, T.A. (1969), *Economics of Labour Force Participation*, Princeton: Princeton University Press.

(9) BREWIS, T.N. (1969), *Regional Economic Policies*, Toronto: The MacMillan Company of Canada Ltd.

(10) CAIN, G.G. (1966), *Married Women in the Labour Force: An Economic Analysis*, Chicago: The University of Chicago Press.

(11) CHAMPERNOWNE, D.G. (1953), A model of income distribution, *Economic Journal*, 54, 318-51.

(12) CHERNICK, S.E. (1966), *Inter-regional Disparities in Income*, Staff Study No. 14, Economic Council of Canada, Ottawa: Information Canada.

(13) CHISWICK, B.R. (1970), An inter-regional analysis of schooling and the skewness of income, in *Education, Income and Human Capital*, edited by HANSEN, W.L., New York: National Bureau of Economic Research.

(14) CHISWICK, B.R. (1971), Earnings inequality and economic development, *Quarterly Journal of Economics*, 85, 21-39.

(15) CHOW, G. (1960), Tests for equality between sets of coefficients in two linear regressions, *Econometrica*, 28, 591-605.

(16) COCHRAN, W.G. (1970), Some effects of errors of measurement on multiple correlation, *Journal of American Statistical Association*, 65, 22-33.

(17) DAVID, M. and MILLER, R. (1970), A naive history of individual incomes in Wisconsin, 1947-59, *Review of Income and Wealth*, No. 1, 79-115.

(18) DENTON, F.T. (1966), *An Analysis of Inter-regional Differences in Manpower Utilization and Earnings*, Staff Study No. 15, Economic Council of Canada, Ottawa: Information Canada.

(19) DERNBURG, T.F., ROSETT, R.N. and WATTS, H.W. (1958), *Studies in Household Economic Behaviour*, New Haven: Yale University Press.

(20) DEVINE, P.J. (1969), Inter-regional variations in the degree of inequality of income distribution: The United Kingdom, 1949-65, *Manchester School of Economic and Social Studies*, No. 2, 141-59.

(21) ELTETO, O. and FRIGYES, E. (1968), New income inequality measures as efficient tools for causal analysis and Planning, *Econometrica*, 36, 383-96.

(22) FISHER, R.A. (1938), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

(23) GOLDBERGER, A.S. (1964), *Econometric Theory*, New York: John Wiley & Sons.

(24) HAINSWORTH, G.B. (1964), The Lorenz Curve as a general tool of economic analysis, *The Economic Record*, 40, 426-41.

(25) HANNA, F.A. (1957), Analysis of inter-state income differentials: Theory and Practice, 113-94 in *Regional Income*, Vol. 21, Studies in Income and Wealth, Princeton: Princeton University Press.

(26) HANOCH, G. (1967), An economic analysis of earnings and schooling, *The Journal of Human Resources*, 2, 310-29.

(27) HANSEN, W.E. (1970), *Education, Income and Human Capital*, Vol. 35, Studies in Income and Wealth, Princeton: Princeton University Press.

(28) HUANG, D.S. (1970), *Regression and Econometric Methods*, New York: John Wiley & Sons.

(29) JOHNSTON, J. (1963), *Econometric Methods*, New York: McGraw-Hill Inc.

(30) KELLY, F.J., BEGGS, D.L. and McNEIL, K.A. (1969), *Research Design in the Behavioral Sciences: Multiple Regression Approach*, Carbondale and Edwardsville: Southern Illinois University Press.

(31) KUZNETS, S. (1963), Distribution of income by size, VIII, in *Quantitative Aspects of the Economic Growth of Nations*, Chicago: University of Chicago Press.

(32) LEVINE, D.B. and SINGER, N.M. (1970), The mathematical relation between the income density function and the measurement of income inequality, *Econometrica*, 38, 324-30.

(33) LONG, C. (1958), *The Labour Force Under Changing Income and Employment*, Princeton: Princeton University Press.

(34) LYDALL, H. (1968), *The Structure of Earnings*, London: Oxford University Press.

(35) MANN, W.E. (1970), *Poverty and Social Policy in Canada*, Toronto: The Copp Clark Publishing Co.

(36) MELICHAR, E. (1965), Least-squares analysis of economic survey data, *American Statistical Association (Proceedings of the Business and Economic Statistics Section)*, 373-85.

(37) MINCER, J. (1958), Investment in human capital and the personal income distribution, *Journal of Political Economy*, 66, 281-302.

(38) MINCER, J. (1962), Labour force participation of married women in *Aspects of Labour Economics*, Princeton: Princeton University Press.

(39) MINCER, J. (1970), The distribution of labour incomes: A survey with special reference to human capital approach, *Journal of Economic Literature*, 8, 1-26.

(40) MORGAN, J.N., DAVID, M.H., COHEN, W.J. and BRAZER, H.E. (1962), *Income and Welfare in the United States*, New York: McGraw-Hill Inc.

(41) MORGAN, J.N. and SONQUIST, J.A. (1964), Problems in the analysis of survey data and a proposal, *Journal of American Statistical Association*, 58, 415-35.

(42) NICHOLSON, N.L. and SAMETZ, Z.W. (1963), Regions of Canada and the regional concept, in *Regional and Resource Planning in Canada*, Edited by KRUEGEL, R.R., SARGENT, F.O., de VOS, A. and PEARSON, N., Toronto: Holt, Rinehart and Winston of Canada Ltd.

(43) OFFICER, L.H. and ANDERSON, P.R. (1969), Labour force participation in Canada, *The Canadian Journal of Economics*, 2, 278-87.

(44) OSTRY, S. (1968), *Provincial Differences in Labour Force Participation*, Ottawa: Statistics Canada.

(45) OSTRY, S. (1968), *The Female Worker in Canada*, Ottawa: Statistics Canada.

(46) PODOLUK, J.R. (1968), *Incomes of Canadians*, Ottawa: Statistics Canada.

(47) PUTNAM, D.F. and PUTNAM, R.G. (1970), *Canada: A Regional Analysis*, J.M. Dent and Sons (Canada) Ltd.

(48) RUTHERFORD, R.S.G. (1955), Income distributions: A new model, *Econometrica*, 23, 277-94.

(49) SCHULTZ, R.R. (1951), On the measurement of income inequality, *American Economic Review*, 41, 107-22.

(50) SITWELL, O.F.G. (1969), Regions as systems: Consideration of a Nova-Scotian case study, in *Perspectives on Regions and Regionalism and other Papers*, edited by CARD, B.Y., Edmonton: University of Alberta Printing Services.

(51) SOLTOW, L. (1969), *Six Papers on the Size Distribution of Wealth and Income*, Vol. 33, Studies in Income and Wealth, Princeton: Princeton University Press.

(52) SPENCER, B.G. and FEATHERSTONE, D.C. (1970), *Married Female Labour Force Participation: A Micro Study*, Special Labour Force Studies, Series B, No. 4, Ottawa: Statistics Canada.

(53) STIGLITZ, J.E. (1969), Distribution of income and wealth among individuals, *Econometrica*, 37, 382-97.

(54) STONE, L.O. (1969), *Migration in Canada: Some Regional Aspects*, Ottawa: Statistics Canada.

(55) SUITS, D.B. (1957), Use of dummy variables in regression equations, *Journal of American Statistical Association*, 52, 548-51.

(56) THUROW, L.C. (1969), *Poverty and Discrimination*, Washington: The Brookings Institution.

(57) TINGERBEN, J. (1957), Welfare economics and income distribution, *American Economic Review (Papers and Proceedings)*, 47, 490-503.

(58) VALLEE, F.G. (1969), Regionalism and ethnicity: The French-Canadian case, in *Perspectives on Regions and Regionalism and other Papers*, edited by CARD, B.Y., Edmonton: University of Alberta Printing Services.

(59) ZEISEL, H. (1968), *Say It With Figures*, New York: Harper and Row Publishers.